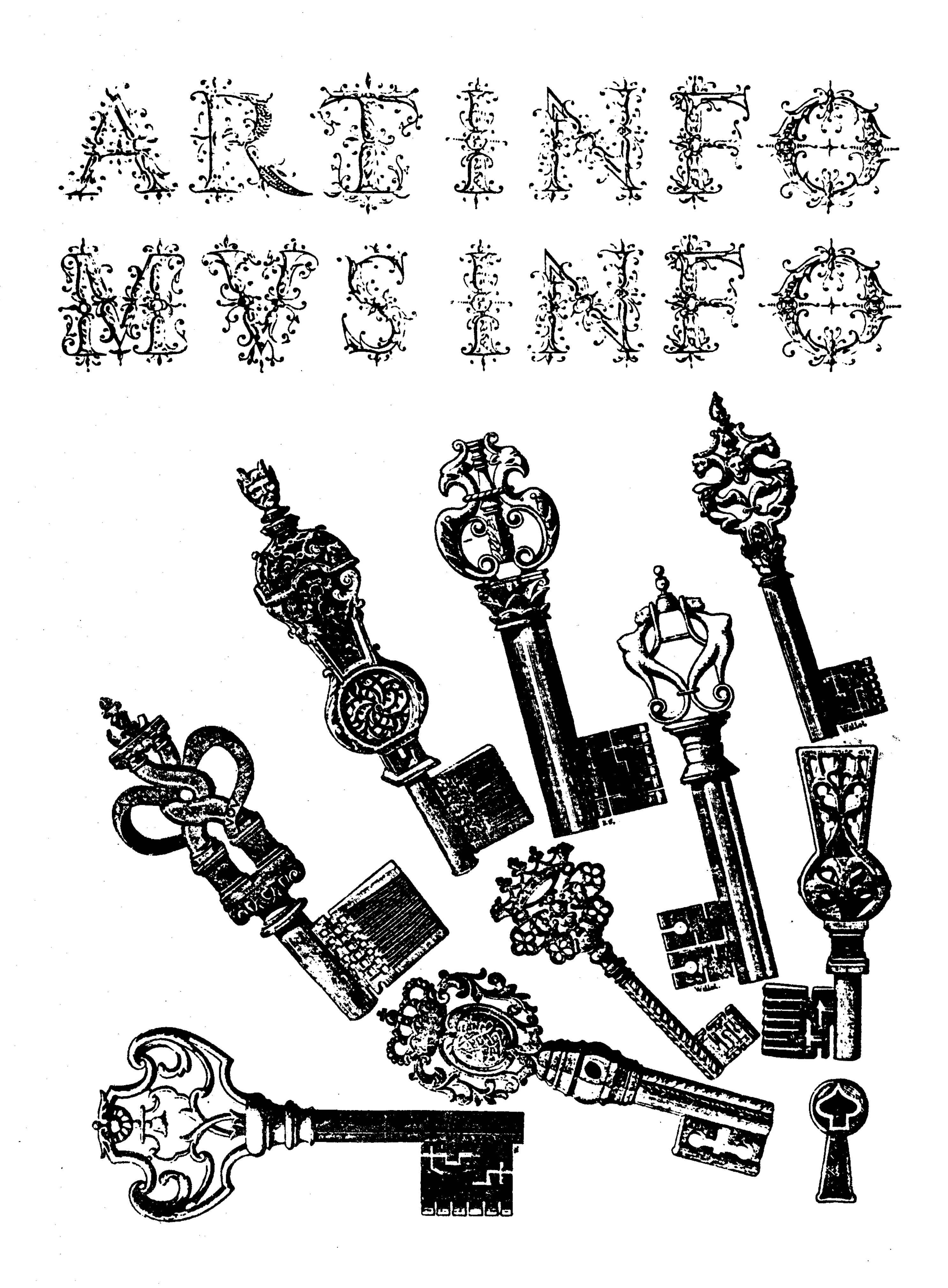

ARTINFO
MVSINFO

Bulletin Technique n° 16
Institut d'Intelligence Artificielle
Université PARIS VIII

ART INVISIBLE, ART HERMÉTIQUE, ART CACHÉ, ART SECRET

Yves LECERF

Conventionnellement nous désignerons par les termes de "phénomènes d'art invisible, hermétique, caché, secret", un ensemble de situations de perception provoquées par des oeuvres d'art lorsque la richesse du complexe des structures à percevoir vient à dépasser pour une raison ou pour une autre les possibilités humaines de perception consciente ou de compréhension effective ; le paradoxe introduit par de telles situations tenant au fait qu'une certaine partie importante de l'oeuvre artistique en cause devra necessairement échapper au spectateur ou à l'auditeur, mais que cependant de cette partie non réellement vue ni perçue, il se dégage un effet artistique supplémentaire qui s'ajoute à celui qu'aurait pu sans cela produire la partie de l'oeuvre consciemment perçue ; si bien que l'impression globale produite par l'oeuvre est différente de celle qu'aurait pu provoquer, présentée seule , la partie de l'oeuvre directement accessible à la conscience.

a) - Art invisible.

Il y a des musiques que l'oreille n'entend pas, des peintures que l'oeil ne voit pas, des images que l'esprit ne perçoit pas : musiques trop lointaines de l'oreille dans leurs écarts de cheminement en direction du grave ou de l'aigu, peintures représentées en détails microscopiques à l'intérieur de toiles peintes de dimensions normales, ou bien images photographiques à défilement trop bref dans le cours d'une oeuvre cinématographique : ces musiques, ces peintures, ces images relèvent d'un art invisible.

b) - Art hermétique.

Il y a des complexités musicales ou graphiques que l'oreille pourrait bien entendre, que l'oeil pourrait bien voir, mais qui dépassent les possibilités instantannées d'analyse et

d'inventaire synthétique de l'esprit humain ; en sorte qu'elles ne sont pas vraiment perçues dans leur totalité et que tout se passe comme si elles n'étaient que partiellement entendues, partiellement vues : elles relèvent d'un art hermétique ou d'un art pour partie hermétique.

c) - Art caché.

Il y a des composantes d'oeuvres musicales ou graphiques qui ne relèvent, au sens de ce qui vient d'être défini, ni d'art invisible, ni d'un art hermétique : car l'oreille peut les entendre sans difficulté, l'oeil peut les voir, et leur complexité n'est absolument pas telle que l'esprit humain soit impuissant à les appréhender ; et composantes qui tout de même ne sont pas perçues ou ne semblent pas perçues, pour la simple raison qu'elles apparaissent dans une sorte de déguisement de l'évidence qui s'apparente à de la prestidigitation ; on pourrait les voir et les percevoir, elles sont visibles et se présentent bien en évidence, et de manière telle cependant que l'attention soit orientée vers la recherche d'autres objets. Ces composantes, visibles et non remarquées, relèvent d'un art caché.

d) - Art secret.

Il y a enfin quelquefois des secrets dans la dynamique de génération de certaines oeuvres d'art ; par exemple dans certaines oeuvres graphiques ou musicales pour lesquelles un système algorithmique vient prêter son appui à l'effort de création , on perçoit comme résultat final une structure apparemment variée et complexe, comportant des régularités dont on ne découvre point la clef si l'auteur ne l'explicite pas . Ces oeuvres, tant que leur clef reste cachée, relèvent d'un processus de création qui contient un secret, autrement dit d'un processus d'art secret.

e) - Un charme supplémentaire indéfinissable.

Art invisible, art hermétique, art caché, art secret : le fait que la perception de certains éléments soit condamnée à demeurer partielle, et le fait que la perception d'autres éléments soit appelée à demeurer en grande partie subconsciente et doive échapper à la mémorisation, tout cela n'exclut absolument pas que le résultat final puisse être tout de même un renforcement de la qualité de l'oeuvre ; aux vertus rationnellement perceptibles et explicables de celle-ci, on voit s'ajouter bien souvent l'enchantement d'une sorte de charme indéfinissable : et bien des écoles hermétistes soutiennent qu'il ne serait se concevoir d'évidence d'art sans l'intervention secrète d'éléments invisibles et point d'affirmation hardie d'une oeuvre sans l'intervention de divers secrets de nombres pour en régir les proportions.

Pour l'intelligence, l'art invisible, l'art hermétique, l'art caché, l'art secret apparaissent comme des énigmes et des défis ; comme sont toujours énigmes et défis , les secrets de tous les trésors cachés. Elleapercevra dans l'art des artifices, et cherchera à travers ces artifices inconnus des clefs dissimulées d'algorithmes à découvrir ; elle concevra l'artefact algorithmique volontariste et la machination création comme des défis lancés aux déterminismes naturels de la création artistique ; et elle percevra l'artefact algorithmique intuitif inspiré et involontaire comme un défi aux capacités des sciences d'analyse logique des structures algorithmiques.

Bulletin Technique n° 16
Institut d'Intelligence Artificielle
Université Paris VIII

L'EFFET "PHRASES CACHÉES"

(méthodes d'analyse stylistique comportant la prise en considération de structures syntaxiques cachées ou virtuelles)

Yves LECERF

GENERALITES : PROBLEMATIQUE DE L'"EFFET PHRASES CACHEES"

Sous le nom générique d'"effet", dans les sciences physiques, il est habituel de désigner des constructions déductives combinant à la fois des observations expérimentales et des hypothèses théoriques, concernant des phénomènes qu'avant leur observation on aurait pu juger assez inattendus, voire même insolites.

Utilisée ici maintenant dans un domaine expérimental assez différent, celui de la linguistique, cette dénomination d'effet qui va être appliquée à ce que nous désignerons ici comme "l'effet phrases cachées" ne nous a pas semblé cependant mal adaptée à son objet ; car les phénomènes mis en cause par "l'effet phrases cachées" concernent une catégorie de chaînes syntaxiques que les linguistes n'ont pas coutume de considérer avec attention : à savoir des chaînes non syntagmatiques, découpées de façon assez quelconque dans des phrases ; et c'est un fait dans une certaine mesure insolite que de telles chaînes puissent avoir des propriétés non aléatoirement quelconques.

On se souviendra en effet que l'effort patient de la linguistique générale puis structurale durant nombre de décennies a tendu précisement au contraire à démontrer que les langues naturelles ne sont pas isotropes au fractionnement ; à démontrer qu'il n'est ni pertinent ni légitime de procéder à des découpages quelconques de la chaîne syntaxique ; à démontrer qu'au contraire un fractionnement légitime du discours ne peut s'effectuer que selon certains segments privilégiés dits "syntagmes" et absolument pas selon un fractionnement quelconque en des chaînes qui incluraient la fin d'un sytagme et le début d'un autre. Pourtant, avec ce que nous appellerons "l'effet phrases cachées" c'est précisement ce genre de chaînes aberrantes que nous allons considérer, les chaînes non syntagmatiques, en vue de décrire leurs propriétés.

Encore, parmi ces chaînes, ferons nous une place toute particulière à une catégorie à priori aberrante : la catégorie des chaînes susceptibles à elles seules de constituer une phrase complète si on les extrait de leur contexte, avec une signification autonome, allant souvent à contresens du document dont on les a extraites ; à ces chaînes, nous donnerons le nom de "phrases cachées".

Les occurences ne manquent pas de ce genre de phrases cachées dans des textes de toutes origines, et l'on voit, et l'on sait à quel point des auteurs, hommes politiques, orateurs, ont pu déplorer souvent la pratique journalistique de la "citation tronquée". Pour illustrer bien toutefois le mécanisme de la phrase cachée, nous mentionnerons quelques exemples :

la phrase : "Jean épouse Jeanne et Christine épouse Christian"	contient la phrase : "Jean épouse Jeanne et Christine"
la phrase : "les voleurs fuyant devant ces chiens policiers totalisaient vingt agressions à main armée"	contient la phrase : "ces chiens policiers totalisaient vingt agressions à main armée"
la phrase : "la robe que j'avais achetée à Adrienne avait complètement rétréci sous la pluie"	contient la phrase : "Adrienne avait complètement rétréci sous la pluie"
la phrase : "le chat de la fermière croque une souris"	contient la phrase : "la fermière croque une souris"

Et il apparaît clairement que des exemples de ce genre pourraient être sans difficulté multipliés à l'infini.

Quelle importance, quelle pertinence, attribuer à cet ensemble de chaînes non syntagmatiques que sont les "phrases cachées"?

C'est toute la problématique de "l'effet phrases cachées" que nous nous proposons d'exposer ici : un point de vue selon lequel ces phrases constitueraient une sorte de face cachée du langage, dont la prise en considération pourrait éclairer la compréhension de phénomènes linguistiques et notamment stylistiques qui sont habituellement considérés comme directement pertinents. Retraçons rapidement l'histoire de cette problématique.

Il n'existe pas à notre connaissance d'étude de psychanalyse du langage qui ait directement pris pour thème les "phrases cachées" ; l'occasion ne s'en est sans doute pas présentée ; mais on peut tenir cependant pour vraisemblable que ces phrases auraient, dans un tel éclairage, été considérées comme porteuses possibles de signification subconsciente.

Toujours est-il que les premiers concours de circonstances qui ont conduit à faire, des "phrases cachées" et des "segments cachés" y afférents, l'objet d'une véritable démarche de réflexion logique et scientifique, datent vraisemblablement des environs des années 1960, avec l'avènement de l'informatique, et la mise en oeuvre d'une première série importante de recherches sur l'analyse algorithmique des langues naturelles dans l'espoir d'applications ultérieures à la traduction automatique des langues.

Parmi les nombreuses difficultés rencontrées dans le déroulement des travaux d'analyse syntaxique automatique des langues, figuraient des phénomènes parasites dus à l'existence des "phrases cachées" et des segments analogues, qui introduisaient souvent des contresens dans les analyses, et conduisaient dans tous les cas à une consommation de temps d'ordinateur.

Par le terme générique de "phrases cachées", nous voulons en effet désigner ici un phénomène d'ensemble, qui porte aussi sur de simples segments pseudosyntaxiques cachés : chaînes nominales cachées, chaînes adjectivales cachées, etc. par exemple :

le segment de phrase :
"le poudrier d'une dame de l'assistance dont le dos est émaillé ..." (I,3)
(d'après Bally)

contient les segments :
"une dame de l'assistance dont le dos est émaillé ..." (I,4)
"l'assistance dont le dos est émaillé..." (I,5)

A propos de ce dernier exemple, on peut se demander si la syntaxe interdit réellement que le dos d'une dame soit émaillé et, sachant que des ordinateurs sont seulement capables d'appliquer strictement des consignes syntaxiques, mais non de tenir compte de la signification des phrases, l'on peut se faire une idée de l'ampleur des difficultés introduites en analyse automatique par l'existence du phénomène de "phrases cachées". Malgré l'importance des efforts de recherche déployés dans le monde depuis plus de dix ans, ces difficultés sont très loin d'avoir pu être surmontées, du point de vue de l'exactitude des réponses données par les machines.

Les phénomènes parasites de "phrases cachées" n'affectaient, dans ces expériences d'analyse syntaxique automatique, pas seulement le résultat du calcul mais aussi, comme déjà dit, sa durée.

Il arrivait souvent, par exemple que tout paraisse fonctionner normalement: les machines ayant reçu des phrases à analyser, avaient abondamment fait clignoter leurs lumières sur les consoles de contrôle, elles avaient fait tourner leurs bandes et leurs disques avec le doux ronronnement d'un bon fonctionnement, elles avaient mené à bien leurs travaux et rendu des réponses exactes. Mais des reconstitutions de l'emploi du temps des unités de calcul, effectuées après coup, montraient que ces organes avaient longuement raisonné sur des "phrases cachées" du type : "la fermière croque une souris" cité plus haut pour conclure d'ailleurs qu'il fallait les rejeter;ceci d'autant plus longuement que pour certaines phrases, les possibilités de fausses pistes étaient vraiment très nombreuses,affectant par exemple simultanément plusieurs groupes de mots : fausses chaînes nominales, fausses chaînes verbales, etc. par exemple :

la phrase :
"le chat de la fermière de mon père croque tous les matins une souris" (I,6)

contient les phrases :
"la fermière de mon père croque tous les matins une souris" (I,7)
"mon père croque tous les matins une souris" (I,8)
"la fermière de mon père croque tous les matins" (I,9)
"mon père croque tous les matins" (I,10)
"le chat de la fermière de mon père croque tous les matins" (I,11)

Le fait qu'une phrase aussi courte et aussi simple puisse contenir déjà rien moins que cinq"phrases cachées" toutes bien construites donne une idée de

l'ampleur que peut atteindre le phénomène et du nombre des possibilités combinatoires examinées par les ordinateurs : ainsi, dans les résultats, rien n'apparaissait des maudites "phrases cachées" mais tout se passait comme si les machines avaient rêvé longtemps à leur sujet.

Pour en terminer avec cet exposé historique, on peut rappeler que les difficultés ainsi rencontrées du fait des "phrases cachées" et de leurs interférences avec les analyses automatiques de phrases normales correspondent à des expériences s'étendant des années 1960 environ jusqu'aujourd'hui puisqu'elles durent encore. Vue sous l'angle spécifique de ces difficultés, et sans approfondissement des concepts en cause, la problématique du phénomène "phrases cachées" se résumerait en quelques mots : un effet parasite, résultant d'un ensemble de petits faits antérieurement considérés comme totalement insignifiants.

Les raisons que nous pouvons éventuellement mettre en avant aujourd'hui, de renverser complètement cette problématique, et d'affirmer que l'effet"phrases cachées" constitue une face cachée du langage, et que, dans la mise en oeuvre du langage, cet "effet phrases cachées" peut impliquer un rôle fonctionnel précis, ces raisons sont à rechercher à la fois sur le plan de la théorie et de l'observation.

Sur le plan de la théorie : s'il s'agit d'éliminer "l'effet phrases cachées" en tant que phénomène parasite, il semble raisonnable de proposer des modèles de calculs d'identification des phrases qui intègrent exclusivement ce phénomène : l'objet de l'algorithme devient alors de calculer à la fois les structures de phrases et les structures de "phrases cachées".

Or nous proposons justement un certain type de modèle algébrique (Algèbre de semi anneau sur un magma) qui présente naturellement cette propriété d'intégrer les calculs de "phrases cachées" (sous forme d'un ensemble de calculs de signification à résultante nulle) au calcul normal. En outre, ce type de représentation algébrique permet d'intégrer naturellement au calcul un certain nombre d'autres structures, ce qui constitue a posteriori une raison supplémentaire de lui attribuer une valeur explicative meilleure que celle de représentations n'intégrant pas les "phrases cachées".

Sur le plan de l'observation : les faits observés à l'occasion d'expériences d'analyse automatique du langage par ordinateur, les difficultés rencontrées du fait des phrases cachées dans le déroulement de ces travaux, tout cela reste insuffisant pour inspirer un renversement de la problématique du phénomène "phrases cachées" : car seuls y sont en cause des ordinateurs, et des méthodes de calcul dont rien ne prouve qu'elles s'apparentent suffisamment à celles de l'intelligence humaine.

Pour affirmer que "l'effet phrases cachées" est une sorte de face cachée du langage, il faut des bases expérimentales d'observation supplémentaires. Il faut montrer que les "phrases cachées" peuvent être impliquées dans des aspects fonctionnels précis du langage, par exemple dans ses aspects stylistiques.

Que la stylistique des "phrases cachées" varie d'un auteur à l'autre, nul n'en doutera à priori et c'est d'ailleurs un fait que l'on peut vérifier. Que la prise en considération de l'ensemble des "phrases cachées" et segments cachés, au lieu de la prise en considéraiton du texte sous sa face directe, puisse constituer une sorte de vision radiographique singulière, éventuellement assez pénétrante en tant qu'instrument d'analyse stylistique, nul n'en doutera a priori, non plus, et c'est d'ailleurs aussi un fait que l'on peut vérifier, sans qu'il soit question

d'exagérer la portée de cette méthode qui restera un simple instrument d'analyse parmi beaucoup d'autres. Tout cela donnerait en tout cas déjà une base d'argumentation pour proposer de renverser la problématique du phénomène "phrases cachées".

Mais que l'on puisse en venir à certaines observations laissant supposer que les "phrases cachées" joueraient quelquefois un rôle fonctionnel direct et directement explicable dans certains effets de style, c'est alors là que des raisons deviendraient vraiment importantes de voir dans les"phrases cachées" une seconde face cachée du langage. Or nous pensons que des apparences très fortes, des présomptions très fortes, existent en faveur d'un rôle direct des "phrases cachées" dans une certaine forme de style poétique : d'une part, les "phrases cachées" et "segments cachés" apparaissent en quantité anormalement élevée, et d'autre part ces"segments cachés" présentent des caractéristiques esthétiques, comme si l'on avait tiré parti de leur existence pour cacher en eux des densités poétiques supplémentaires.

A cet ensemble d'observations et de considérations théoriques que nous venons d'énumérer, il nous semble au total justifié de donner le nom proposé en début de ce chapitre, de problématique de "l'effet phrases cachées".

DES POEMES CACHES DANS DES "PHRASES CACHEES" DANS DES POEMES

Il existe au moins deux raisons évidentes de présumer que les textes poétiques en général peuvent constituer un terrain d'étude privilégié pour le phénomène des "phrases cachées".

L'une des raisons tient aux usages sémantiques de la littérature poétique, qui font à l'insolite une assez bonne place, surtout à l'insolite dans des présentations à moitié hermétiques du fait d'un certain caractère caché ; et les quelques exemples de mécanismes de "phrases cachées" cités plus haut ont montré que ces phrases pouvaient avoir des significations quelquefois étranges.

La seconde de ces raisons tient, au contraire, non plus à la sémantique mais à la forme des textes de littérature poétique : ces textes sont précisément entre tous, les seuls à mettre en évidence à l'intérieur de leurs propres phrases une segmentation supplémentaire complètement anormale, qui est celle de la versification. Cette désinvolture formelle des"phrases cachées" à l'égard de la syntaxe des phrases normales qui les hébergent, syntaxe que les"phrases cachées" prennent délibérément en biais ou à contre sens, nous la retrouvons dans la désinvolture de versification de maints poèmes :

"Douces colonnes, aux
Chapeaux garnis de jour" (Valéry)

"Parmi les marronniers, parmi les
Lilas blancs, les lilas violets" (Moréas)

Quant à l'organisation méthodique des études que l'on peut envisager de conduire sur le rôle des"phrases cachées"dans la littérature poétique c'est très naturellement qu'elle trouve son point d'appui et son point de départ dans une recherche exhaustive des "phrases cachées" dans les textes poétiques.

Pour dresser ainsi la liste de ces "phrases cachées", il pourrait sembler normal et naturel d'utiliser des ordinateurs puisqu'aussi bien leur aptitude particulière à détecter de tels segments est justement à l'origine de notre démarche ; et l'on pourrait espérer que leur puissance de calcul permette de traiter rapidement une large collection de textes.

Cependant, le parti que nous devrons adopter consistera précisément, au contraire à nous contenter de techniques artisanales, appliquées à un certain nombre de textes très courts choisis comme échantillons : les "phrases cachées" seront recherchées si l'on peut dire "à la main", sans moyen mécanique ou électronique auxiliaire, mais en reproduisant artisanalement et dans la mesure du nécessaire, les opérations qu'aurait effectuées un ordinateur. Cette renonciation à l'emploi de moyens mécaniques s'explique par plusieurs raisons, dont la principale tient à l'état des "programmes" actuellement existants dans le monde pour l'analyse syntaxique automatique de textes en langue française : il s'agit seulement de programmes expérimentaux, dotés chacun d'un vocabulaire et d'une syntaxe de dimensions très restreintes ; ces programmes sont utiles pour observer expérimentalement à petite échelle certains phénomènes, mais demeurent totalement impropres à un quelconque traitement de masse de textes français ; il est donc hors de question de mettre en oeuvre une expérience de recherche massive de "phrases cachées" par ordinateur.

Ajoutons que la perspective de réaliser, pour l'avenir, une expérience massive de ce genre demeure d'autant plus éloignée que le corpus à traiter est celui des textes poétiques, c'est-à-dire le corpus de textes dont la syntaxe présente les pires difficultés imaginables du point de vue d'un dispositif d'analyse grammaticale automatique.

Du reste, le fait que la recherche "artisanale", manuelle, des "phrases cachées" demeure aujourd'hui et pour longtemps la seule modalité pratiquement envisageable n'entraîne pas nécessairement que cette procédure soit dépourvue d'avantages et d'efficacité. Le choix d'un nombre suffisant d'échantillons, certes assez courts, peut nous donner certainement une première approximation des phénomènes étudiés, un peu comme un "sondage" peut donner, à partir d'échantillons, une approximation d'une situation statistique. Pour ce qui concerne le dépouillement de chaque texte, on s'efforcera d'imiter la démarche des ordinateurs, en observant que leur efficacité à découvrir des "phrases cachées" tient tout simplement au caractère exhaustif des découpages de segments usuellement commandés par les programmes d'analyse : il y aura donc lieu de procéder tout simplement à un découpage systématique, combinatoire et exhaustif des textes, suivi d'un examen également systématique de chaque segment. Certains segments révèleront la présence de "phrases cachées", et l'on en dressera l'inventaire.

Un avantage du caractère artisanal du procédé est de permettre de consigner dans ce même inventaire diverses autres particularités notées pour chacun des segments rencontrés, particularités par exemple de versification qui relèvent de critères non syntaxiques et n'auraient pas été détectables par des programmes d'analyse automatique ; ou bien détection de certaines catégories très particulières de "phrases cachées" non mécaniquement inventoriables en tant que n'ayant pas la structure de phrases bien constituées.

Dans la catégorie des segments acceptables au titre de "phrases cachées", il semble en effet raisonnable de faire une place à certains segments comparables dans leur structure à ces phrases courtes et lapidaires que sont les "titres de livres" comme par exemple "Pour qui sonne le glas", "Vipère au poing","Sous le soleil de Satan". De tels segments sont indiscutablement placés"entre un point et un point"dans les textes imprimés, comme des phrases normales. Les textes poétiques peuvent en

contenir, comme on verra plus loin, une assez importante proportion ; par exemple, trois en un seul vers d'Eluard :

"... Pointe à frayeur, fruit en guenilles, amour rapace ..."

Nous verrons, à l'occasion d'exemples donnés plus loin, que la signification de certains de ces segments après leur découpage va parfois complètement à l'encontre de la signification du texte dont ils sont extraits ; ce qui est la preuve de leur qualité de "phrases cachées" autonomes, de la même façon que "la fermière croque une souris" constituait un contresens par rapport à la signification principale "le chat de la fermière croque une souris" dans un exemple déjà cité.

Ces opérations de recensement de l'ensemble des "phrases cachées" et "segments cachés" une fois menées à bien, il reste à en tirer parti dans des observations stylistiques.

Du point de vue qualitatif, il faudra apprécier l'ensemble des segments ainsi sélectionnés, et cela implique sans doute une certaine part d'arbitraire. Toutefois une mention particulière doit être donnée des cas éventuels où certains segments-vers en forme classique (alexandrins), etc. - pourraient être détectés dans les "phrases cachées" car l'identification de structures de ce genre ne peut être considérée comme relativement objective. Mention particulière également doit être donnée des "segments titre de livre" dont il vient d'être parlé plus haut, car eux aussi se prêtent à une appréciation relativement extérieure au domaine de la poésie, donc relativement objective : la question de savoir si un segment donné pourrait faire ou non un bon titre de livre met en jeu une intuition relativement mécanique, elle appelle facilement une réponse par oui ou par non, avec de bonnes concordances d'opinion entre plusieurs sujets interrogés. C'est un test que nous utiliserons largement dans la suite de cette étude.

Du point de vue quantitatif, il sera bon de comparer des textes poétiques et des textes en prose ; c'est ce que nous ferons pour mettre en évidence la différence de densité en "segments cachés" qui sépare ces deux modes d'expression stylistique.

DES MICRO-POEMES EN FORME DE NOMS DE POEMES OU NOMS DE LIVRES OU NOMS DE TABLEAUX CACHES DANS DES POEMES

Il a déjà été fait allusion plus haut à cette catégorie très particulière de phrases que sont les titre de poèmes, les titres de livres, les noms de tableaux, et l'on a fait observer que de tels segments pouvaient être à bon droit considérés comme des phrases, certes très courtes. Il semble justifié de les prendre en compte dans des inventaires de "phrases cachées".

C'est avant tout un style, riche en oppositions et violent dans ses images, qui caractérise ce genre de segments que sont les titres ; et il y a en effet certainement une gageure à essayer de donner un pouvoir d'évocation à des segments aussi courts. Pourtant les titres ont ce pouvoir : "Sang et lumière" ; "Une fille de Saragosse" ; "Tendre et violente Elisabeth" ; "Par qui le scandale" ; "Soudain l'été dernier" etc.

Si l'on omet dans une évaluation de leur longueur les mots incolores tels que "une" "de" "la" etc. on observe que les titres sont dans l'ensemble construits sur deux ou trois notions sémantiques principales ; par exemple trois pour "La

grande peur dans la montagne", pour "Baigneuse du clair ou sombre",pour "Blason des fleurs et des fruits", pour "Le quatrième poème visible" ; et par exemple deux pour "Mort en fraude", "Les anges noirs", "Les violons parfois" ; et même quelquefois une seule notion comme dans "La peau", "La peste", "Olivia", "Le mépris". En tous cas, les noms de livres sont des segments courts et doivent, pour des raisons mnémotechniques, être tels.

Si, inversant le problème, nous posons la question de savoir ce que pourrait être un genre poétique de poèmes très courts, soumis à la contrainte d'un maximum de trois notions, il nous semble qu'un tel genre aurait beaucoup de parenté avec la littérature des titres de livres. L'extrême brièveté du genre imposerait que l'on recherche là aussi des compensations soit par le moyen d'oppositions très contrastées, soit par le choix d'images violentes ou riches en évocation, soit encore par le recours à des découpages insolites. Et quant à l'idée même de poème très court, elle n'est pas si absurde, puisque la littérature japonaise connaît avec le Haiku un genre de poème à cinq ou six notions.

Nous pensons donc que des segments en forme de titres de poèmes, titres de livres, noms de tableaux, trouvés dans des poèmes ordinaires au hasard des textes, présenteraient le double intérêt d'être d'une part une certaine forme de"phrases cachées" et d'autre part une certaine forme de"micropoèmes cachés".

A leur sujet, nous énonçons une affirmation : la fréquence des"phrases cachées" en forme de titres de poèmes,titres de livres, noms de tableaux, rencontrés dans les textes poétiques dépasse nettement celle atteinte dans les autres textes.

Et sur ce même sujet, nous ajouterons que cette singularité est un des traits qui contribuent le plus à créer dans les textes poétiques la situation de "multiple fractionnabilité" dont la description fait l'objet de la présente étude. En effet, les noms de poèmes ou de livres ont une structure, on l'a vu, très concise et ramassée : "Croisade sans croix", "Pleure ô mon pays bien aimé", "Le baiser au lépreux". Par ailleurs, ils ne constituent pas nécessairement une chaîne d'apparence syntagmatique normale : "Les violons parfois", "Par qui le scandale" ; "Dans un mois, dans un an", "Soudain l'été dernier", etc.

La vérification d'une propriété de fractionnabilité selon des segments de ce genre laisse sans doute quelque marge d'appréciation à la subjectivité. Soit l'exemple d'un quatrain d'Eluard (A la flamme des fouets) :

> "Métal qui nuit, métal de jour, étoile au nid,
> Pointe à frayeur, fruit en guenilles, amour rapace,
> Porte couteau, souillure vaine, lampe inondée,
> Souhaits d'amour, fruits de dégoût, glaces prostituées ..."

Il nous semble quant à nous que tous les segments constitutifs du texte (segments intercalés entre virgules, tous sans exception) pourraient faire de remarquablement bons titres de romans ou poèmes. De la même façon une lecture du "Cantique des colonnes" de Valéry permet de relever des : "hymnes candides", "servantes sans genoux", "filles des nombres d'or", "un dieu couleur de miel" auxquels nous prêtons la même qualité ; et de même une lecture du poème "A la nue accablante tu" de Mallarmé nous donne des segments tels que : "de basalte et de laves", "échos esclaves", "trompe sans vertu", "sépulcral naufrage", "perdition haute", et ainsi de suite. Mais pour tous ces segments, nous ne sommes pas en mesure de citer de précédent défini de livre, poème, film, tableau, portant précisément ces noms.

Ce que l'on peut affirmer, relativement à de tels précédents, c'est que sans aucun doute des emprunts sont constamment faits au corpus des textes poétiques existants par des auteurs cherchant des titres pour des oeuvres nouvellement composées.

"Le sel de la terre"	(Gide)	origine : nouveau testament
"Par qui le scandale"	(Hougron)	origine : nouveau testament
"Bonjour tristesse"	(Sagan)	origine : Eluard

Sans doute serait-il souhaitable de disposer d'un inventaire exhaustif des emprunts de ce genre ou même des coïncidences involontaires ; mais de toute manière, l'ensemble des segments de poèmes est beaucoup plus vaste que celui des titres déjà attribués à d'autres oeuvres ; et un inventaire de ce genre ne suffirait donc même pas à apprécier le phénomène dans toute son ampleur; il faudrait de toute manière avoir recours à l'appréciation stylistique de segments comme titres possibles.

Puisqu'aussi bien le titre "Bonjour tristesse" vient d'être mentionné comme extrait d'un poème d'Eluard, considérons ce poème : il s'intitule "A peine défigurée" et son texte est le suivant :

"Adieu tristesse
Bonjour tristesse
Tu es inscrite dans les lignes du plafond
Tu es inscrite dans les yeux que j'aime
Tu n'es pas tout à fait la misère
Car les lèvres les plus pauvres te dénoncent
Par un sourire
Bonjour tristesse
Amour des corps aimables
Puissance de l'amour
Dont l'amabilité surgit
Comme un monstre sans corps
Tête désappointée
Tristesse beau visage"

En dehors du segment "Bonjour tristesse" effectivement choisi comme titre d'un roman, on observe que bon nombre de chaînes de mots ont le style des titres de livres : "Adieu tristesse", "les lignes du plafond", "les yeux que j'aime", "la misère", "les lèvres les plus pauvres", "les plus pauvres te dénoncent", "Par un sourire", "Amour des corps aimables", "Puissance de l'amour", "monstre sans corps", "Tête désappointée" ; et même aussi "tristesse bonjour", "tristesse tu es", "sourire bonjour", "des corps aimable puissance", bien que ces dernières chaînes soient un peu moins bonnes ; en outre, bon nombre des vers du poème semblent acceptables en qualité spécifique de titres de poèmes mais peut-être moins facilement de titres de livres : c'est le cas par exemple du segment "tristesse beau visage" dont le caractère irrationnel s'accomoderait mieux d'un poème que d'un roman.

Il est important de le faire observer, les significations évoquées par certains de ces segments se différencient très nettement de celles constituant la trame du poème ; par exemple "des corps aimables puissance" est probablement un solécisme et un contresens, car puissance se réfère à "puissance de l'amour" et non aux corps aimables ; par exemple encore "les plus pauvres te dénoncent" est un solécisme et évoque une image peut-être voisine du texte mais tout de même assez distincte ; par exemple encore "sourire bonjour" est un solécisme et sa signification n'a guère de rapport avec celle du passage dont on le tire, et ainsi de suite.

Une caractéristique de la"phrase cachée" consiste en effet précisément dans ceci qu'elle possède une autonomie propre de signification ; par exemple la signification de "la fermière croque une souris" s'écartait nettement de celle de la phrase d'origine "le chat de la fermière croque une souris". En retrouvant cette autonomie chez les segments titres de livres, nous obtenons une confirmation supplémentaire de leur aptitude à constituer de véritables "phrases cachées" et non plus seulement des fragments de phrases normales.

Ce phénomène d'autonomie de signification des segments découpés est observable dans bon nombre de poèmes d'Eluard. Par exemple dans le quatrain :

"Une ou plusieurs
L'azur couché sur l'orage
La neige sur les oiseaux
Les bruits de la peur dans les bois revêches".

Chacun des vers peut être considéré comme un titre de livre possible et même assez brillant : "une ou plusieurs", "l'azur couché sur l'orage", "la neige sur les oiseaux","les bruits de la peur dans les bois revêches" ; mais en outre d'autres découpages du texte suggèrent des titres à contresens : "sur l'orage la neige", "sur les oiseaux les bruits de la peur" dont l'existence confirme le statut d'autonomie de signification des segments : ce sont bien des "phrases cachées".

Toujours dans ce même quatrain, notons que l'énumération du sens et du contre-sens n'épuise pas complètement la liste des "segments titres de livres": des chaînes telles que "couché sur l'orage" ou "la peur dans les bois" ou "les bruits de la peur" etc. constituent des titres acceptables, de même encore que "l'azur couché", "les bois revêches", "sur l'orage" etc. Certes, elles ne forment pas des contre-sens caractérisés, mais on ne saurait exiger que toutes les "phrases cachées" aillent à contre-sens de la signification principale. Observons enfin que le total des segments notés pour ce simple quatrain devient tout de même considérable si l'on se réfère aux dimensions modestes du texte dont ils sont extraits : leurs longueurs cumulées dépassent largement le double de la longueur du texte.

Les traits observés seraient-ils particuliers aux poèmes d'Eluard? Il n'en est rien bien entendu. Prenons au hasard un poème de Baudelaire, par exemple la très célèbre "Invitation au voyage":

"Mon enfant, ma soeur
Songe à la douceur
D'aller là-bas vivre ensemble!
Aimer à loisir
Aimer et mourir
Au pays qui te ressemble!
Les soleils mouillés
... etc."

Ici encore le texte est constitué presque en totalité de segments susceptibles de servir de titres à des livres : "Enfant ma soeur", "songe", "douceur d'aller", "songe à la douceur", "là-bas vivre ensemble", "aimer à loisir", "ensemble aimer", "aimer et mourir", "au pays qui te ressemble", "mourir au pays", "à loisir aimer", "les soleils mouillés" etc. Et ici encore, l'autonomie de la signification est évidente pour des titres tels que "mourir au pays" lequel évoque immédiatement un souci très répandu de dormir son dernier sommeil nulle part ailleurs qu'au cimetière du village natal, donc le contraire de l'invitation au voyage, le contraire du sens principal. Et ici encore, la densité du texte en segments dans le style de titres de livres apparaît comme très riche.

Un inventaire que l'on voudrait vraiment exhaustif de tous les segments de textes susceptibles ainsi de servir de noms à d'autres oeuvres, poèmes, romans, tableaux conduirait parfois à retenir la presque totalité des segments, c'est à dire même

pour des phrases très courtes un nombre assez considérable de chaînes. Prenons l'exemple d'une phrase courte choisie au hasard dans le roman "Dedans" d'Hélène Cixous, au milieu de la première page du livre:

"Dans mon jardin d'enfer les mots sont mes fous".

On peut découper dans cette phrase plusieurs segments très typiques du style des titres de livres en général : "jardin d'enfer", titre qui suggère un récit fantastique à la Poe ; "les mots", titre réel d'un ouvrage de Sartre ; "d'enfer les mots", qui fait songer aux aphorismes d'Esope sur le langage ; "les mots sont", titre qui nous renvoie à la caverne de Platon et à la querelle des universaux. La diversité même de ces significations montre une fois de plus que les "segments titres de livres" sont de véritables phrases cachées, aussi distinctes du sens principal que notre "phrase cachée" type "la fermière croque une souris" citée plus haut pouvait l'être de sa phrase mère "le chat de la fermière croque une souris". Toujours dans la même phrase d'Hélène Cixous, d'autres titres peuvent encore être trouvés : "dans mon jardin", évocateur de joies bucoliques ; "les mots sont mes fous", hermétique. "Enfer" qui est presque un titre de Dante ; "Fous", pour citer les segments les plus évidents seulement. Un découpage exhaustif de la phrase conduit à découvrir encore quelques segments acceptables supplémentaires mais qui ne sont que des variantes des précédents ; par exemple : "jardin" et "mon jardin", distincts du segment "Dans mon jardin" déjà cité, avec d'ailleurs de nettes possibilités de divergence de signification ; par exemple "Mots" au lieu de "les mots" déjà cité ; par exemple "d'enfer" au lieu de "enfer" déjà cité , etc.

Au total, on parvient sans doute à découvrir quelques segments vraiment inacceptables, tels que "d'enfer les " ou "dans mon", segments dont un découpage combinatoire révèle l'existence et qui ne peuvent en aucun cas être rattachés au style des titres de livres ; mais il est remarquable d'observer que ces segments constituent plutôt l'exception. Le test de la recherche de phrases cachées par découpage de la phrase initiale d'Hélène Cixous se révèle donc extrêmement positif.

Mais cet exemple, tiré à la différence des précédents d'un texte en prose, pose le problème de la distinction entre prose poétique et prose tout court.

"L'EFFET PHRASES CACHEES" ET LA PROSE

C'est en l'opposant à"la prose" que l'on peut donner de la littérature poétique les définitions les plus simples, dans des formulations qui toutefois recèlent bien des dangers. C'est en tout cas dans "la prose", indiscutablement, que réside l'ultime élément de preuve, dont nous avons lieu de faire état pour démontrer le caractère singulier des phénomènes de "microphrases cachées"décelés dans de nombreux textes poétiques : cette singularité particulière, il faut montrer que les textes "en prose" ne l'ont pas.

C'est, il est vrai, à dessein que nous avons choisi comme dernier exemple de texte poétique, un texte en prose, le roman "Dedans" d'Hélène Cixous : il y a des textes en prose qui relèvent du statut de la littérature poétique, il y a des poèmes en prose,il y a de la prose poétique ; bref, il y a parfois des poèmes cachés dans de la prose.

Compte tenu de la complexité de cette situation, une discussion devra nécessairement suivre l'examen du ou des contre-exemples choisis dans "de la prose qui soit vraiment de la prose" ; mais nous suggérons qu'il soit procédé d'abord à cet examen des textes eux-mêmes, en commençant par une lettre adressée le 6 mars 1697 à M. Ménage, par un de nos plus admirables prosateurs français, Mme de Lafayette :

> "J'attends présentement Clélie avec toute l'impatience que j'ai accoutumé d'avoir pour elle ; mais cette impatience est encore augmentée par les louanges que vous lui donnez. Quand vous louez quelque chose pour être bien écrit, il faut s'en fier à vous ; car, pour l'ordinaire, vous ne louez guère que ce qui mérite d'être loué. Je pourrais tirer de là une conséquence qui me serait assez avantageuse ; car vous m'avez tant louée en français et en italien que je serais en droit de croire que j'aurais quelque sorte de mérite. Mais ma vanité ne me mène pas si loin et j'attribue les louanges que vous m'avez données à l'amitié que vous avez pour moi, sans y chercher d'autre cause"

Il est selon notre sentiment impossible de découper le moindre titre de livre un peu incisif ou percutant de l'ensemble de ces lignes. A la rigueur, les quatre premiers mots "J'attends présentement Clélie", pourraient-ils rappeler l'ennui voulu d'un "En attendant Godot", mais de bien loin ; sans doute aussi, certains mots isolés pourraient-ils constituer des titres d'ouvrages, mais d'une formulation bien terne : "amitié" "l'amitié" "impatience" "mérite" "louanges" ; à la rigueur aussi quelques groupes de mots tels que "l'impatience que j'ai accoutumé d'avoir" ou bien "quelque sorte de mérite", mis en valeur par un découpage insolite et par l'insolite de découper une banalité ; mais on voit bien que ces quelques possibilités de titres n'ont de justification que par antiphrase, autrement dit parce qu'il ne s'agit pas de titres normaux ; et encore sont-ils bien insignifiants. Le décompte total des "titres de livres" possibles est donc bien proche, pour un découpage dans ce texte, d'un état néant.

Pour s'en convaincre mieux encore, l'on peut soumettre une des phrases, par exemple la troisième, à un fractionnement exhaustif :

> "Quand vous/vous louez/quand vous louez/louez quelque chose/vous louez quelque chose/louez quelque chose pour/quelque chose pour être/quelque chose pour être bien/pour être/pour être bien/pour être bien écrit/être bien écrit/bien écrit,il/écrit,il faut/écrit il faut s'en/il faut s'en fier/faut s'en fier/faut s'en fier à/ s'en fier à vous/fier à vous/".
> "Quand/vous/louez/quelque/chose/pour/être/bien/écrit/il/faut/s'en/ fier/à/vous/."

Il n'y a pas de segment acceptable pour un titre dans cet ensemble sauf peut-être, à la rigueur, certains mots isolés : par exemple "être". C'est peu.

Le "contre exemple" ayant ainsi apporté confirmation de son caractère différent des "exemples" cités avant lui, exemples de poèmes dont on pouvait extraire bon nombre de segments en forme de noms de poèmes, noms de livres, noms de tableaux, il nous reste à discuter de cette distinction qui justement existe entre "prose poétique" et "prose tout court".

Une théorie séduisante consisterait à poser en principe que la propriété faisant l'objet de notre étude peut justement constituer un critère de cette distinction : mériterait le nom de "prose poétique" toute prose dans laquelle il serait possible de discerner de nombreux fragments ayant l'éclat d'un "minipoème" en forme de nom de poème, et ceci même de préférence, en sorte que le découpage soit continu.

Quant à l'incidence de "l'effet phrases cachées" dans les textes en prose, il ne faudrait pas se hâter de conclure qu'elle est nécessairement nulle : on peut penser dans la mesure où il ne s'agit pas de prose poétique, ni de poèmes en prose, que cette incidence prend d'autres formes, qui varient selon les auteurs.

Sans doute avons-nous pu constater que le style de Madame de Lafayette présente une assez faible densité de "phrases cachées" allant à contre-découpage et à contre-sens du texte, comme en "phrases cachées" du type "titres de livres". Son cas peut être rapproché de celui d'un autre prosateur féminin, auteur célèbre en littérature enfantine, la Comtesse de Ségur ; et l'on conçoit du reste assez bien qu'il en soit ainsi pour cette dernière ; l'existence de "phrases cachées" à signification ésotérique ou à contresens ne serait certainement pas de nature à favoriser la facilité de lecture pour un jeune public. Pour ce qui concerne les propriétés de fractionnabilité et segmentabilité de la prose de la Comtesse de Ségur, on aura lieu d'observer qu'elle reste curieusement élevée, mais cette segmentabilité conduit à des chaînes dont les significations confirment chaque fois celles des phrases dont elles sont extraites,et dont les structures confirment celles normalement assimilables aux phrases ; en outre chaque chaîne exprime le plus souvent une affirmation qui pourrait être considérée comme vraie isolément, par référence au sens général du texte.

Exemple de texte :

> "Après donc m'avoir tenu les propos les plus humiliants, auxquels je ne répondais qu'en me léchant les lèvres et en lui tournant le dos, elle prit son bâton et se mit à me battre si cruellement que je finis par perdre patience et que je lui lançai trois ruades, dont la première lui cassa le nez et deux dents, la seconde lui brisa le poignet, et la troisième l'attrapa à l'estomac et la jeta par terre"
> (Comtesse de Ségur : "Les mémoires d'un âne")

Segments remarquables :
"je ne répondais qu'en me léchant les lèvres et en lui tournant le dos"
"elle prit un bâton et se mit à me battre"
"je finis par perdre patience"
"je lui lançai trois ruades"
"la première lui cassa le nez et deux dents"
"la seconde lui brisa le poignet"
"la troisième l'attaqua à l'estomac et la jeta par terre"

Ces segments remarquables sont grosso modo tous des syntagmes du texte, mais non pas des phrases ni même des propositions indépendantes coordonnées : et cependant la situation est bien telle que si des parties de texte échappent à l'attention des jeunes lecteurs, les parties remarquables qui restent comprises sont interprétables comme des phrases et leur sens confirme celui du texte.

UN MODELE ALGEBRIQUE POSSIBLE POUR LA REPRESENTATION DE "L'EFFET PHRASES CACHEES"

Parmi les formalismes que l'on peut proposer pour représenter à la fois une phrase et ses "phrases cachées", l'un des plus naturels est celui de la fonction Π, "produit contracté dans un magmoïde". On trouvera une description détaillée de ce formalisme dans l'article "Méthodes de calcul pour une analyse informatique des structures syntaxiques cachées de textes poétiques"("Effet phrases cachées").

Indiquons brièvement ici ce que nous entendons par magmoïde : c'est une algèbre de semi anneau A définie sur un magma E, avec opération extérieure de multiplication par des pseudo scalaires appartenant à un semi anneau α défini sur un second magma ε.

Indiquons d'autre part ce que nous entendons par produit contracté Π. La définition récursive de cette fonction est la suivante :

$$\Pi(a_1 a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n) \; \Pi(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

$$\Pi(a_j) = a_j \quad \text{pour tout } a_j \in E$$

formule dans laquelle les produits et les sommes doivent s'interpréter en termes d'opérations du semi anneau E ; dans laquelle $(\overset{i}{\rightarrow} a_1 a_2 a_3 \ldots a_n)$ désigne la chaîne $(a_1 a_2 \ldots a_n)$ limitée à ses i premiers termes;et dans laquelle $(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$ désigne la chaîne $(a_1 a_2 \ldots a_n)$ limitée à ses n-i derniers termes.

Expliquons pourquoi ce formalisme est proposé.

Une des manières les plus simples que l'on puisse trouver de représenter une famille de structures de phrases ou de fragments de phrases, consiste à recourir à des structures d'arbres articulées autour des éléments (monèmes, mots) qui constituent ces phrases ou fragments ; et la structure algébrique de magma permet de représenter un arbre par un terme monome produit non associatif ; une structure de semi anneau (algébrique) A sur un magma E permettra donc de représenter des familles d'arbres dans un formalisme polynomial.

Pour arriver à faire des tris de sous-catégories et de sous familles d'arbres, en sorte à parvenir à pondérer les énumérations d'arbres par des coefficients susceptibles de s'annuler, il est nécessaire d'introduire une opération de multiplication extérieure, par exemple par un scalaire ; mais on voit bien que la multiplication par des pseudo scalaires relevant eux aussi d'une algèbre de semi anneau α sur magma ε ne présente pas plus de difficulté formelle, car le calcul des pseudo scalaires prend appui sur les formalismes de toutes façons déjà nécessaires pour les éléments du magma E déjà existant ; et le fait de disposer de pseudo scalaires (à table de multiplication arbitraire) donne des possibilités de tris bien supérieures à celles des scalaires simples pour sélectionner des sous ensembles d'arbres dans E.

Dans le cas particulièrement simple où E est un magma libre et ε un magma fini, la structure de magmoïde se rapproche beaucoup d'une description de grammaire context-free, les éléments de E figurant des arbres et leurs coefficients étant égaux à la catégorie syntaxique (non terminale) du syntagme correspondant à chaque arbre.

Quant au produit contracté Π, ce serait une fonction égale à la somme de toutes les structures d'arbres possible, s'il n'était le rôle des coefficients pseudo-scalaires nuls qui font que Π se réduit à un sous ensemble particulier ; en l'absence de coefficients nuls, Π regroupe vraiment toutes les structures : c'est une fonction universelle de compilation.

Le fait de choisir une représentation de phrase par des arbres peut sembler de nature à introduire une certaine limitation mais il faut bien noter que les pseudo scalaires extérieurs sont utilisables en nombre quelconque (sur des magmas auxiliaires $\varepsilon_1 \varepsilon_2 \varepsilon_3 \ldots \varepsilon_p$ etc.) et que ces pseudo scalaires sont potentiellement aussi des arbres;autrement dit, l'élément de base de cette structure n'est pas l'arbre simple mais le produit cartésien de plusieurs arbres.

On en trouvera dans le document déjà cité en référence une description beaucoup plus détaillée, ainsi qu'une étude des propriétés principales de ce système formel, notamment sur l'aptitude à représenter des contraintes complexes faisant intervenir de façon combinée et imbriquée des contraintes relevant à la fois de plusieurs grammaires context-free différentes ; notamment aussi son aptitude à représenter des langages complètement non context-free, tels que le langage écho x-x.

On trouvera dans un autre article, intitulé "Méthodes de calcul pour une évaluation de l'effet versification cachée dans la versification" des détails sur la possibilité de représenter, par ces moyens, des contraintes introduisant des relations complexes entre des structures de versification et des structures syntaxiques. Le champ d'application de ce type de calculs s'étend notamment à la caractérisation de structures versifiées éventuellement incluses dans des phrases cachées quelconques, elles sont données par la fonction Π, sous réserve d'ajuster convenablement les coefficients.

DES PERCEPTIONS SUBCONSCIENTES DE FRAGMENTS DE PHRASES CACHEES DANS LA PERCEPTION CONSCIENTE DES PHRASES ; DES CHAINES FAUSSEMENT SYNTAXIQUES CACHEES DANS DES RAISONNEMENTS SYNTAXIQUES CORRECTS

S'il n'est pas invraisemblable que la qualité poétique des parties constituantes d'un texte, prises selon de multiples fractionnements entrecroisés, contribue dans un sens favorable à la qualité d'ensemble du texte dans son entier, il reste que l'on est en droit de s'interroger et de rester perplexe sur ce qui concerne la possibilité effective pour l'auditeur ou le lecteur du poème de percevoir à la fois tant de "beautés" dans tant de fragments.

Des phrases cachées dans des phrases ; des poèmes cachés dans des poèmes ; des "micro-poèmes" en forme de noms de poèmes, noms de livres, noms de tableaux, noms de romans, cachés dans des poèmes ; des poèmes cachés dans de la prose ; des versifications insolites cachées dans des phrases cachées ; que perçoit-on vraiment de tout cela?

Certes, les fragments d'un texte, et tous ses fragments, sont dans la situation de se trouver énoncés lorsque le texte est prononcé (et dans la situation d'être parcourus par l'oeil lorsque le texte est lu). Un auditeur frappé d'une amnésie telle qu'il ne pourrait mémoriser que les quatre ou cinq derniers mots prononcés devant lui, serait en cette qualité d'amnésique tout particulièrement apte à percevoir les fragments d'un poème en tant qu'unités distinctes. Mais pour un auditeur normal, le fait que les fragments soient énoncés ne suffit pas à démontrer que ces fragments sont perçus en tant que tels.

Quant au caractère à exiger d'une telle perception, il semble raisonnable de se contenter éventuellement d'une perception subconsciente. S'agissant de sentiment esthétique, une perception consciente n'est en effet pas à priori indispensable.

Il ne serait donc pas indifférent de pouvoir apprécier la fréquence usuelle de perception subconsciente, l'importance usuelle de perception subconsciente des éléments poétiques cachés dans un poème ou cachés dans de la prose. S'il pouvait être établi que la perception subconsciente de ces "éléments cachés" est usuelle fréquente, importante, alors l'on serait sans doute conduit à formuler l'hypothèse que ces éléments poétiques cachés jouent directement un rôle fonctionnel déterminé dans le "charme poétique". S'il pouvait au contraire être établi que la perception subconsciente des éléments poétiques cachés est toujours nulle, alors, l'on serait conduit à formuler l'hypothèse que ces éléments cachés n'ont pas de rôle fonctionnel direct dans le "charme poétique".

Nous ne disposons guère d'éléments factuels susceptibles de permettre l'appréciation de cette éventuelle perception subconsciente ; du moins, guère d'éléments pour cette perception subconsciente par des gens ; car au contraire, nous disposons d'éléments factuels très nombreux relatifs à des machines, relatifs à des situations qu'à tort ou à raison nous estimons comparables, et où la perception est le fait d'une machine et non pas d'une personne.

Nous l'avons dit plus haut, l'actuelle problématique de "l'effet phrases cachées" et du rôle des "phrases cachées" dans la littérature poétique tire sa source d'observations effectuées à propos d'expériences d'analyse automatique de textes par ordinateur. Dans ces expériences portant sur des phrases ordinaires, on décelait des phénomènes parasites de "phrases cachées" : par exemple, la phrase déjà citée : "le chat de la fermière croque une souris" se révélait contenir la phrase cachée : "la fermière croque une souris", laquelle phrase cachée ne manquait pas d'être perçue et analysée par les programmes pour être ensuite rejetée comme non compatible avec les mots immédiatement précédents. Le travail ainsi effectué pour l'analyse des "phrases cachées" pouvait être cause de pertes notables en temps machine, ceci d'autant plus que le phénomène commençait au niveau de simples segments faussement syntaxiques, puis s'étendait de là à des combinaisons de ces segments, lesquelles combinaisons pouvaient se révéler assez nombreuses.

Ainsi, l'un des plus difficiles problèmes, parmi tous ceux auxquels s'est trouvée confrontée la recherche en analyse automatique de textes, a consisté et consiste encore à essayer en quelque sorte d'empêcher la "perception" des "phrases cachées" par les machines ; au moins à limiter les conséquences de cette perception en temps perdu à raisonner sur ces pseudo-phrases.

Il est particulièrement instructif de constater que, malgré plus de dix années d'efforts déployés dans les équipes de recherche de diverses universités du monde, il n'a nulle part été possible d'éliminer les phénomènes parasites liés à la prise en considération par les machines des "phrases cachées". Il s'agissait et s'agit encore pourtant uniquement d'un problème de programmation et de syntaxe à la fois : les sous programmes chargés de regrouper les mots en syntagmes et en phrases sont généralement très bien séparés de ceux remplissant d'autres fonctions, on les appelle programmes de "parsing", et c'est de leur organisation que relève l'élimination des phrases ou syntagmes parasites. De nombreuses améliorations ont été proposées pour ces techniques de parsing, d'autres les sont encore ; certaines des modifications que l'on peut apporter à ces programmes tendent par exemple à accélérer les raisonnements sur les "phrases cachées" ou bien à rendre ces raisonnements partiellement implicites au lieu d'explicites. Mais toutes les tentatives que l'on a pu faire pour sauter purement et simplement l'étape du raisonnement sur les "phrases cachées" se sont traduites immanquablement par l'omission d'une partie des réponses correctes : mélangées aux "phrases cachées", une partie des phrases correctes, des solutions exactes, étaient éliminées sans examen ; et les analyses effectuées par les machines, si elles avaient l'avantage de se dérouler plus rapidement, le compensaient par l'inconvénient de donner moins souvent de bonnes réponses.

La prise en considération en tant que telles des pseudo phrases cachées constitue-t-elle une étape nécessaire au raisonnement syntaxique correct? Autrement dit, l'énonciation, au moins implicite, de l'erreur pour justement l'écarter constitue-t-elle une étape du raisonnement correct en analyse syntaxique automatique? L'expérience semble montrer qu'il en est ainsi. Pour des cas particuliers, le raisonnement permet même de l'affirmer. Pour certaines phrases en effet, empêcher l'unité de calcul de formuler certaines hypothèses fausses puis de les éliminer, c'est adopter un dispositif obligeant cette même unité de calcul à se tromper dans l'interprétation d'autres phrases pour lesquelles ces mêmes hypothèses se trouveraient au contraire

justement fondées. En voici des exemples :

le matin de la pluie serait bienvenu (VI,1)
la venue de la pluie serait bienvenue (VI,2)
le matin de la pluie dont je vous ai parlé ... (VI,3)
le matin de la pluie précéda l'émeute (VI,4)
le matin de la pluie avait précédé le matin de l'émeute (VI,5)

Dans les phrases (VI,1 à 5) ci-dessus, la chaîne "de la pluie" est tantôt complément de nom (VI,2,3 et 5) tantôt chaîne nominale sujet (VI,1 et 4), avec du reste une marge résiduelle non négligeable d'ambiguité. La chaîne "le matin" est tantôt nominale sujet (3,5) tantôt chaîne complément de temps (1,4). Quant à la question de savoir si la chaîne "le matin de la pluie" a une existence syntagmatique, sa réponse est tantôt oui (3,5) tantôt non (1,4).

L'analyse des phrases (VI,1 à 5) ne peut se faire rigoureusement qu'en essayant chaque fois pour "le matin de la pluie" les deux possibilités. La réussite dans une hypothèse ne dispense pas d'examiner l'autre, comme le montre la phrase tout à fait ambiguë :

"le matin de l'orage serait bienvenu" (VI,10).

L'histoire des techniques de parsing est riche d'exemples de phrases choisies par hasard comme phrases type et dont les auteurs découvraient après coup, avec la plus grande surprise, qu'elles comportaient une ambiguité (exemples : "le pilote ferme la porte", "they are flying planes" etc...).

Ainsi, la réponse correcte à un problème d'analyse syntaxique d'une phrase en langue naturelle inclut un double contenu : premièrement, la liste des interprétations syntaxiquement acceptables ; secondement, la démonstration qu'il n'y a pas d'autre interprétation syntaxiquement acceptable. Cette dernière démonstration peut souvent résulter pour certaines de ses parties de déductions implicites ; mais l'examen explicite de certaines chaînes virtuellement syntagmatiques (telle : "le matin de la pluie") en vue de les rejeter est un examen inévitable : l'énoncé de l'erreur pour la réfuter apparaît comme une étape nécessaire du raisonnement correct.

C'est là une des différences qui séparent le problème théorique de l'analyse des phrases du problème de la génération des phrases..Dans un processus génératif, par exemple pour la génération d'un langage de Chomsky, on ne considère que des chaînes syntagmatiques correctes :

S → NP + VP
NP → Art + N
VP → V + NP ... etc

Dans un processus d'analyse au contraire, on doit considérer toutes ces chaînes syntagmatiques correctes, plus un certain nombre de chaînes virtuellement syntagmatiques pour les éliminer. Ainsi, l'énoncé de l'erreur constitue une étape nécessaire du raisonnement correct en analyse.

Des raisonnements syntaxiques que nous venons de faire ci-dessus, et qui sont confirmés par l'expérience de nombreuses tentatives d'analyse de textes par ordinateurs, il convient de reporter maintenant les conclusions au problème dont nous étions partis, le problème de la perception subconsciente éventuelle de segments de textes poétiques : et nous voyons qu'il en résulte des présomptions importantes en faveur d'une réponse positive.

Certes le comportement de machines programmées ne peut en général pas être considéré comme un précédent indicatif de ce que serait le comportement d'êtres humains accomplissant des tâches comparables ; mais les contraintes que les problèmes de

machines ont fait découvrir ici sont des contraintes purement syntaxiques et logiques apparemment inhérentes à la nature du langage humain : de telles contraintes s'appliquent aux personnes aussi bien qu'aux machines.

Il semble donc probable que les "phrases cachées", phrases pseudo-syntaxiques et segments pseudo-syntaxiques cachés dans les phrases constituent une sorte d'écho intérieur du discours dans le discours, un écho qui serait perçu de façon inconsciente au cours du processus même de reconnaissance du langage. Et l'on ne saurait exclure qu'il puisse en résulter pour certains textes un effet esthétique.

Bulletin Technique n° 17
Institut d'Intelligence Artificielle
Université PARIS VIII

MÉTHODES DE CALCUL POUR UNE ANALYSE INFORMATIQUE DES STRUCTURES SYNTAXIQUES CACHÉES OU VIRTUELLES DE TEXTES POÉTIQUES — (effet "phrases cachées")

Yves LECERF

- C H A P I T R E 0 -

OBJET : CALCULS POUR "L'EFFET PHRASES CACHEES"

Les méthodes de calcul qui vont être décrites ont été développées à propos de l'étude de ce que nous avons appelé par ailleurs "l'effet phrases cachées". Cet "effet" (le mot étant pris ici dans l'acception que lui donnent les physiciens pour désigner un fait expérimental insolite) concerne les fragments que l'on peut obtenir dans un texte en procédant à des découpages complètement aléatoires et quelconques, sans en respecter les unités syntaxiques, et bien entendu sans s'obliger à aller d'un point initial de phrase à un point final de phrase.

Le fait expérimental insolite consiste premièrement dans l'observation qu'un nombre relativement important de segments peuvent avoir une cohérence apparente ainsi que des apparences de signification allant souvent à contresens de celle du texte, en sorte que l'on en vient à associer au texte un ensemble de "segments cachés cohérents" qui lui est propre ; et le fait expérimental insolite consiste en second lieu dans l'observation de la possibilité d'une analyse stylistique de cet ensemble de phrases, ensemble qui peut constituer un indicateur plus sensible, plus accessible à des méthodes d'investigations formelles que le texte lui-même.

On trouvera dans nos deux études "L'effet phrases cachées" (Méthodes d'analyse stylistique comportant la prise en considération de structures syntaxiques cachées ou virtuelles) (dans ce même numéro d'ARTINFO/MUSINFO) et "Des poèmes cachés dans des poèmes" (éléments pour une analyse informatique de structures cachées en littérature poetique) (à paraître dans Poétique) des précisions sur ce qu'est "l'effet phrases cachées" et de nombreux exemples de sa manifestation dans divers textes.

Conceptuellement, l'existence expérimentale des phénomènes que nous venons de désigner sous le nom "effet phrases cachées" et leur possibilité d'exploitation en critique stylistique peut être rattachée à l'observation naïve, pure et simple des textes : on aperçoit des segments, on les découpe, on les compare à ceux d'autres textes.

Mais en pratique, l'idée de découper des textes de façon aléatoire ne vient pas naturellement, à moins que des raisons théoriques n'aient laissé prévoir l'apparition d'un ensemble ayant une existence pertinente, comme résultat d'une telle opération.Pour ce qui concerne notre démarche, c'est l'emploi d'un certain type de calcul syntaxique, avec utilisation d'une algèbre de semi anneau sur un magma, qui l'a suggérée ; car dans une telle formalisation des calculs syntaxiques, on voit apparaître avec un rôle extrêmement important un certain outil mathématique multiforme dénommé "produit contracté" ; ce produit contracté apporte notamment une formalisation directe des propriétés de "compilabilité" d'une phrase,et équivaut à un programme universel de compilation.

Or ce même produit contracté, considéré dans des systèmes de pondération légèrement différents, associe justement à toute phrase du texte, l'ensemble de ses "segments cachés cohérents" : autrement dit, l'emploi d'un formalisme de semi-anneau sur un magma conduit tout droit à des concepts exprimant l'effet "phrases cachées".

C'est ce formalisme que nous proposons de décrire ici.

Les lignes qui vont suivre, et qui concernent la définition et les propriétés élémentaires d'une algèbre de semi-anneau sur un magma, paraîtront peut être un peu techniques.

Ce genre de développements peut en général, lorqu'on utilise des structures mathématiques usuelles telles que des anneaux, des corps, des groupes, des espaces vectoriels, etc. pour des applications linguistiques, être évité en faisant report à la littérature mathématique qui énonce les axiomes et propriétés des dites structures.

Mais justement, il n'existe pas à notre connaissance de publication ayant énuméré les propriétés des algèbres de semi-anneau sur un magma : le sujet est donc en principe entièrement non traité, et la rigueur exige que nous énoncions complètement les axiomes de définition de l'algèbre que nous voulons utiliser.

Ce travail préliminaire est nécessaire pour pouvoir passer à une seconde étape qui sera la définition d'un outil permettant une manipulation algébrique de collections de structures et qui sera le "produit contracté". Mais pour commencer il convient d'en rester au "produit simple", non associatif, dans un semi-anneau que nous allons définir sur un magma : ce sera l'objet d'un chapitre I.

La série des chapitres 2 - 3 - 4 qui fera suite sera tout entière relative à l'outil que nous voulions justement en venir à définir et qui est le "produit contracté".

Dans un chapitre 2 on donnera d'abord (2,1) la définition de ce produit: c'est une fonction qui a les propriétés algébriques d'un produit, et qui cependant va constituer un instrument de manipulation de collections de structures, car le "produit contracté" représente toutes les structures associables à une séquence de "mots" donnée.

Dans un paragraphe (2,2) on montrera que le produit contracté est une fonction récursive et on donnera (2,3) la formule de sa définition par récurrence, on étudiera quelques unes de ces propriétés (2,4) et on évoquera le problème de sa représentation par les programmes (2,5)

Dans un chapitre 3 on montrera que le produit contracté coincide avec le fonction de compilation de toute une famille de langages, incluant notamment les context-free.

Dans un chapitre 4, on examinera l'utilisation du produit contracté pour la prise en compte de "l'effet phrases cachées".

- C H A P I T R E 1 -

DEFINITIONS D'UNE ALGEBRE DE SEMI-ANNEAU SUR MAGMA : MAGMOIDE

Définitions

On dira qu'un ensemble E est doté d'une structure de semi anneau sur un magma si cet ensemble est pourvu de deux lois de composition respectivement dénommées addition et multiplication, satisfaisant aux axiomes suivants :

- a) pour la loi de multiplication E est un magma, doté d'un élément neutre (élément unité I) ainsi que d'un zéro (tous produits par un zéro égaux à zéro)

- b) pour la loi d'addition, E est un semi groupe abélien avec élément neutre 0, qui est le zéro de la multiplication

- c) la multiplication est distributive à gauche et à droite par rapport à l'addition.

Loi externe de multiplication par un scalaire : on peut doter en outre l'algèbre de semi anneau définie ci-dessus d'une loi externe de multiplication par un scalaire, dénommée également "Produit" vérifiant les axiomes de :

- distributivité de ce produit par rapport à l'addition scalaire
- associativité pour les multiplications scalaires
- distributivité de ce produit par rapport à l'addition du semi-anneau
- neutralité pour l'élément unité I.x = x

1.2 - Propriétés élémentaires de la multiplication du semi-anneau.

On peut rappeler que cette multiplication définit une structure de magma sur E, c'est à dire :

1.2.a- La multiplication n'est pas associative.

Autrement dit le produit x(y,z) n'est pas généralement égal (pour tous x,y,et z) au produit (x,y)z : autrement dit encore : $x,y,z \rightarrow x(yz) \neq (xy)z$ cette non associativité est particulièrement adaptée à refléter la non associativité des processus d'identifications syntaxiques, par exemple :

Le premier = x
Vide = y
La coupe = z

(le premier vide) la coupe ≠ le premier (vide la coupe)

x y z ≠ x y z

c'est à dire :

$$(x\ y)\ z \neq x\ (y\ z)$$

1.2.b - La multiplication n'est pas commutative

Autrement dit : le produit (x,y) n'est pas nécessairement équivalent à (y,x), autrement dit encore :

$$(x,y) \rightarrow xy \neq yx$$

Cette non commutativité est adaptée à refléter la non commutativité des processus d'identification syntaxique, par exemple :

grand = x
homme = y

grand homme ≠ homme grand

x y ≠ y x

c'est à dire :
$xy \neq yx$

1.2.c - La multiplication admet un élément zéro (élément permis à droite et à gauche)

Cet élément est tel que pour tout x

$$0.x = x.0 = 0$$

Cet élément est précisément aussi, comme on le verra plus loin, un zéro (élément neutre) pour l'addition : et cette propriété sera utilisée pour la manipulation d'ensembles de structures syntaxiques, lorsqu'il s'agira d'éliminer certaines d'entre elles.

1.2.d - La multiplication admet un élément neutre

Cet élément est donc I tel que pour tout x

$$I.x = x.I = x$$

Cet élément neutre est unique (à noter que zéro n'est pas élément neutre pour la multiplication).

Un tel élément n'a guère de représentation concrète dans le langage (sauf par exemple le blanc qui sépare les mots, mais on n'en voit pas à première vue l'utilisation).

En fait pourtant, il constituera un instrument très utile de calcul pour permettre de créer des expressions contenant à la fois un produit et ses constituants.

Exemple : pour (x,y) : on ajoute I à chacun des constituants et on fait le produit

$$(x + I) (y + I) = xy + x + y$$

Pour des expressions contenant un nombre de facteurs supérieur à deux, on fait intervenir des procédés un peu plus compliqués mais sensiblement de même nature, qui seront décrits plus loin et l'élément unité y joue bien entendu un rôle.

1.2.e - La multiplication est toujours définie

Cette propriété exprime que pour tout couple x et y, d'éléments de E le produit xy existe : elle est une conséquence implicite de la notion d'opération.

Du point de vue des applications à la syntaxe, cette propriété peut surprendre : mais il faut bien voir que l'on dispose de la latitude de faire égal à zéro le résultat de la multiplication, car

$$xy = 0$$

entre dans la catégorie des opérations définies.

1.3 - Propriétés élémentaires de l'addition du semi-anneau et sa distributivité par rapport à la multiplication

Rappelons que cette addition définit une structure de monoïde sur E, il en résulte que :

1.3.a - L'addition est associative

$$(u + v) + w = u + (v + w)$$

La conséquence étant qu'une somme d'un nombre quelconque de termes se note entre deux parenthèses seulement .

$$\begin{aligned}(u + v + w + x + y) &= ((((u + v) + w) + x) + y) \\ &= (u + (v + (w + (x + y)))) \\ &= \text{etc.}\end{aligned}$$

1.3.b - L'addition est commutative (monoïde abélien)

$$u + v = v + u$$

Il en résulte en utilisant l'associativité que l'on peut écrire dans un ordre quelconque les termes d'une somme.

$$u + v + w + x + y = y + x + w + v + u$$
$$= y + u + x + w + v$$
$$= \text{etc.}$$

1.3.c - L'addition admet un élément zéro (élément neutre)

$u + 0 = 0 + u = u$ pour tout u E

On démontre facilement que l'élément neutre est unique : donc zéro est le seul élément neutre pour l'addition, à ce propos il y a lieu de rappeler que cet élément zéro (neutre) pour l'addition coïncide avec le zéro (élément permis à gauche et à droite) pour la multiplication, c'est à dire :

$$0\,x = x\,0 = 0$$

et que la notion d'élément permis se distingue de celle d'élément neutre.

1.3.d - Distributivité à gauche par rapport à la multiplication

$$u(x + y) = ux + uy$$

1.3.e - Distributivité à droite par rapport à la multiplication

$$(x + y)u = xu + yu$$

1.3.f - Conséquence pour les formalismes habituels du calcul

Bon nombre d'identités remarquables du calcul algébrique sont transposables pour des calculs dans le semi-anneau, sous réserve de bien préciser l'ordre des opérations de produit.

Exemple : $(x + y)(u + v) = xu + xv + yu + yv$
$((x + y)z)(u + v) = (xz)u + (xz)v + (yz)u + (yz)v$

Il faut toutefois noter que l'addition n'admet pas d'opposé, la multiplication pas d'inverse, ce qui ôte la possibilité de certaines formes de calcul faisant intervenir explicitement ou implicitement ces notions.

L'addition,dotée de son élément zéro, va constituer pour notre objectif (calcul sur des ensembles de structures linguistiques) le moyen de manipuler des collections de segments de phrases, le zéro permettant d'éliminer de ces collections les segments que l'on ne souhaite pas y retenir : on fait en sorte que ces segments soient des produits de valeur nulle, ou inclus dans des produits de valeur nulle.

1.4 - Propriétés élémentaires relatives à la loi externe de multiplication, par des scalaires d'un anneau A

Soient les scalaires en lettres grecques, et les éléments de E en lettres latines.

1.4.a - Distributivité par rapport aux scalaires

$$(\alpha + \beta)x = \alpha x + \beta x$$

1.4.b - Distributivité par rapport aux éléments

$$\alpha(x + y) = \alpha x + \alpha y$$

1.4.c - Neutralité pour l'élément unité de l'anneau scalaire A

$$1.x = x$$

1.4.d - Multiplication par le zéro de l'anneau scalaire A

$$0.x = 0$$
$$0.1 = 0$$

1.4.e - Associativité relative au produit des scalaires

$$\alpha(\beta x) = (\alpha\beta)x$$

1.4.f - Associativité généralisée des scalaires

$$(\alpha x)(\beta y) = (\alpha\beta)(xy)$$

L'ensemble de ces propriétés va nous permettre d'effectuer des calculs de comptages sur des structures, une collection de structures pouvant être représentée sous forme polynomiale.

1.4.g - Cas particulier

Un cas particulier remarquable est celui ou le corps K se limite à un seul élément : la multiplication scalaire ne joue plus de rôle, tous les scalaires ayant même poids ; mais une conséquence est que l'addition du semi-anneau devient alors idempotente.

$$\forall x \quad E \qquad x + x = x$$

Ce type particulier d'algèbre est utile pour faire des comptages de formes de structures, lorsque l'on n'a plus besoin de tenir compte du nombre de fois où une même forme de structure est représentée.

1.5 - Magmoïde

On dira qu'un ensemble E est doté d'une structure de Magmoïde avec opération extérieure de multiplication par des pseudo-scalaires d'un ensemble ε si

- E est doté d'une structure de semi-anneau sur un magma

- ε également

- il existe une opération extérieure de multiplication scalaire des éléments E par des éléments scalaires pris dans ε ; opération satisfaisant aux axiomes suivants : distributivité par rapport aux éléments de E ; distributivité par rapport aux éléments aussi de ε ; associativité dans un produit de deux facteurs, des scalaires entre eux et des éléments de E entre eux, chacun des deux produits restant non associatif en lui-même ; neutralité pour l'élément unité de ε, produit nul pour l'élément zéro de ε ou pour celui de E.

- CHAPITRE 2 -

DEFINITION DU PRODUIT CONTRACTE SUR LE MAGMA E

2.1 - Produit contracté de degré n

La loi de multiplication n'étant pas associative, des expressions telles que :

$$a(bc) \quad \text{et} \quad (ab)c$$

ne sont pas équivalentes. Nous désignerons du nom de Π_3 (a,b,c) (produit contracté de degré 3) la somme

$$\Pi_3(a,b,c) = a(bc) + (ab)c$$

l'indice 3 sert à rappeler qu'il y a trois variables dans la définition de Π_3.

De même des expressions telles que :

$$(ab)(cd) \quad \text{et} \quad (a(bc))d \quad \text{et} \quad a((bc)d) \quad \text{et} \quad ((ab)c)d \quad \text{et} \quad a(b(cd))$$

ne sont généralement pas égales et nous désignerons du nom de $\Pi_4(a,b,c,d)$ (produit contracté du degré 4 de a,b,c,d) la somme

$$\Pi_4(a,b,c,d) = (ab)(cd) + (a(bc))d + a((bc)d) + ((ab)c)d + a((bc)d)$$

où l'indice 4 vient rappeler que Π_4 est définie comme une fonction de 4 variables, dans l'ordre précité.

D'une manière générale, nous désignerons du nom de $\Pi_n(a_1,a_2, \ldots ,a_n)$ la somme de tous les produits possibles faisant intervenir $(a_1, a_2, \ldots ,a_n)$ dans cet ordre mais avec des parenthetisations différentes.

Pour que la fonction Π_n ait un sens avec n = 1 ou avec n = 2, nous poserons

$\Pi_1(a) = a$ quel que soit a
$\Pi_2(a,b) = ab$ quels que soient a et b

en observant du reste que ces définitions conventionnelles ne contreviennent pas à la définition générale donnée pour Π_n.

2.2 - Définition récursive des fonctions Π_n

2.2.a - Définition préliminaire

Opérateurs $(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)$ et $(a_1 a_2 \ldots a_n \overset{j}{\leftarrow})$. On désignera par $(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)$ la chaîne composée des i premiers éléments de la suite $a_1 a_2 \ldots a_n$. L'opérateur $\overset{i}{\rightarrow}$ n'a de sens que si $i \leq n$; il définit une application du monoïde E^* sur lui même. De même on désignera par $(a_1 a_2 \ldots a_n \overset{j}{\leftarrow})$ la chaîne composée des j derniers éléments de la suite $a_1 a_2 \ldots a_n$; l'opérateur $\overset{j}{\leftarrow}$ n'a de sens que si $j \leq n$; il définit une application du monoïde E^* sur lui même.

Les opérateurs $(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)$ et $(a_1 a_2 \ldots a_n \overset{j}{\leftarrow})$ sont récursifs, et aisément représentables dans les langages de programmation usuels.

2.2.b - Définition de Π_n à partir des Π_k de degré inférieur

La formule de la définition est la suivante :

$$\Pi_n(a_1 a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi_i(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)\, \Pi_{n-i}(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

2.2.c - Vérification de l'équivalence des définitions pour les degrés 2,3,4

Degré 2

éléments de calcul : $\Pi_1(a) = a$ et $\Pi_1(b) = b$

La formule donnée ci-dessus pour Π_n avec n quelconque devient :

$$\Pi_2(ab) = \Pi_1(a)\,\Pi_1(b)$$

et le résultat $\Pi_2(ab) = ab$ est bien celui déjà donné pour Π_2

Degré 3

La formule donnée plus haut pour Π_n avec n quelconque s'applique avec sommation pour i = 1 et i = 2, soit

$$\Pi_3(abc) = \Pi_1(a)M_2(bc) + \Pi_2(ab)M_1(c)$$

ce qui est bien la définition déjà donnée, car cette somme vaut bien (ab)c + a(bc)

Degré 4

La formule donnée plus haut pour Π_n avec n quelconque

$$\Pi_n(a_1a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi_i(\overset{i}{\rightarrow} a_1a_2 \ldots a_n)\, \Pi_{n-i}(a_1a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

s'applique par sommation pour i = 1, i = 2 et i = 3, soit :

$$\Pi_4(abcd) = \Pi_1(a)M_3(bcd) + \Pi_2(ab)M_2(cd) + \Pi_3(abc)M_1(d)$$

c'est à dire

$$\begin{aligned} {}_4(abcd) &= a(b(cd) + (bc)d) + (ab)(cd) + (a(bc) + (ab)c)d \\ &= a(b(cd)) + a((bc)d) + (ab)(cd) + (a(bc))d + ((ab)c)d \end{aligned}$$

ce qui est bien la définition déjà donnée.

2.2.d - Equivalence des deux définitions pour n quelconque

Montrons que $\Pi_n(a_1a_2 \ldots a_n)$ défini par la formule de récurrence se compose de tous les produits possibles, faisant intervenir $a_1a_2 \ldots a_n$ dans cet ordre avec des parenthétisations différentes.

- a) Tous les termes de Π_n en écriture développée sont des produits faisant intervenir $a_1a_2 \ldots a_n$ dans cet ordre. (En effet si la propriété est vraie pour les degrés inférieurs à n, la structure de la formule fait qu'elle est vraie pour Π_n)

- b) Ces termes sont tous différents. (En effet si la propriété est vraie pour les degrés inférieurs, la structure de la formule fait qu'elle est vraie pour n)

- c) Tous les termes que l'on peut former à partir de $a_1a_2 \ldots a_n$ dans cet ordre, avec n'importe quelle parenthétisation, sont représentés. (En effet : soit un terme quelconque ($Prod_n$ $a_1a_2 \ldots a_n$) avec une certaine parenthétisation ou le mot "$prod_n$" signifie qu'une certaine parenthétisation interne existe, mais que l'on ne l'a pas précisée graphiquement ; et supposons la propriété vraie pour tous les Π_k de

degrés $k \leq n$; alors le terme quelconque choisi comporte une dernière parenthétisation qui le décompose en produit de deux facteurs $(Prod_i\ a_1 a_2 \ldots a_n)$ et $(Prod_{n-i}\ a_1 a_2 \ldots a_n)$ mais justement $\Pi_i (a_1 a_2 \ldots a_n)$ contient $(Prod_i\ a_1 a_2 \ldots a_n)$ puisque degré $i < n$ et $(\Pi_{n-i} a_1 a_2 \ldots a_n)$ contient $(Prod_{n-i} a_1 a_2 \ldots a_n)$ puisque degré $n-i < n$ donc Π_n qui contient $\Pi_i \Pi_{n-i}$ contient $(Prod_i a_1 a_2 \ldots a_n)(Prod_{n-i} a_1 a_2 \ldots a_n)$ donc contient le terme donné $(Prod_n a_1 a_2 \ldots a_n)$

2.3 - Produit contracté, définition générale

Jusqu'à présent nous avons utilisé non pas une fonction produit contracté mais une famille de fonctions Π_n produits contractés de degré n (d'ordre n)

Considérons la fonction universelle $\Pi(n, a_1 a_2 \ldots a_n)$ de tous les Π_n : la mention de la variable n y joue visiblement un rôle redondant, puisqu'il suffit de compter les autres variables. Donc $\Pi(a_1 a_2 \ldots a_n)$ constituera une notation générale du produit contracté, équivalent à l'ensemble des notations particulières où le degré est mentionné, et la définition s'écrit :

$$\Pi(a_1 a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi(\overset{i}{\rightarrow}\ a_1 a_2 \ldots a_n)\ \Pi(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

2.4 - Propriété de distributivité de produit contracté par rapport à l'addition

2.4.a - Somme sur un terme

$$\Pi(a_1 a_2 \ldots (a_i' + a_i'' + a_i''') \ldots a_n) = \Pi(a_1 a_2 \ldots a_i' \ldots a_n) + \Pi(a_1 a_2 \ldots a_i'' \ldots a_n) + \Pi(a_1 a_2 \ldots a_i''' \ldots a_n)$$

en effet, chacun des deux membres peut être montré équivalent à la somme de tous les produits faisant intervenir les facteurs $a_1 a_2 \ldots a_i \ldots a_n$ dans cet ordre, avec toutes les parenthétisations possibles, la variable a_i prenant les valeurs a_i' ou a_i'' ou a_i''' et les valeurs des autres variables étant des valeurs $a_1 a_2 \ldots a_{i-1} a_{i+1} \ldots a_n$ déterminées.

Il y a lieu d'observer que cette règle calque la règle de distributivité d'un produit simple dans un anneau ordinaire où l'on a :

$$a_1 a_2 \ldots (a_i' + a_i'' + a_i''') \ldots a_n = a_1 a_2 \ldots a_i' \ldots a_n + a_1 a_2 \ldots a_i'' \ldots a_n + a_1 a_2 \ldots a_i''' \ldots a_n$$

2.4.b - Somme sur plusieurs termes

La règle sera calquée sur celle d'un produit par une somme dans un anneau ordinaire, par exemple :

$\Pi((a + b)(c + d)) = \Pi ac + \Pi ad + \Pi bc + \Pi bd$

La procédure générale est la suivante : on considère $\Pi(a_1 a_2 \ldots a_n)$ comme une fonction des variables $(a_1 a_2 \ldots a_n)$ dont chaque facteur peut prendre au choix les valeurs des termes de la somme qui lui correspond ; et l'on fait la somme des Π obtenus pour toutes les combinaisons des valeurs de ces variables.

2.5 - Programmation de Π, produit contracté

Soit un langage de programmation

- autorisant les définitions "récursives" (appel d'une fonction ou d'une procédure par elle-même)
- comportant des procédures ou fonctions représentatives des calculs "somme" et "produit" du magmoïde
- comportant des procédures ou fonctions représentatives des opérateurs $(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)$ et $(a_1 a_2 \ldots a_n \overset{i}{\leftarrow})$
- comportant des procédures ou fonctions représentatives de l'opérateur sommation $\sum f(i)g(i)$, pour i allant de i = 1 à i = n-1
- comportant une procédure ou fonction représentative du "degré" d'une chaîne (nombre de lettres)

Alors la définition

$$\Pi(a_1 a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n)\, \Pi(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

peut se programmer en une seule instruction dans un tel langage (en prenant soin d'exprimer que pour des chaînes de degré 1, Π est égal à la chaîne elle-même ; et que pour des chaînes de degré supérieur à 1 Π est donné par la formule de sommation).

- C H A P I T R E 3 -

OPERATION EXTERIEURE DE PRODUIT PAR UN SCALAIRE OU PSEUDO-SCALAIRE

LE PRODUIT CONTRACTE COMME FONCTION UNIVERSELLE DE COMPILATION

3.1 - Généralités : scalaires, pseudo-scalaires

L'opération extérieure de multiplication par un scalaire ordinaire (élément d'un corps K ou d'un anneau A) s'introduit pour les structures de semi-anneau sur un magma, qui font l'objet de notre étude, de manière analogue à la façon dont on introduit pour la théorie des vecteurs la multiplication extérieure par un scalaire.

Mais il est évident que disposant de la notion de semi-anneau sur un magma, nous l'exploitons beaucoup plus naturellement et surtout beaucoup plus complètement en définissant l'opérateur extérieur comme la multiplication par un "pseudo-scalaire", lequel sera un élément non pas d'un corps, non pas d'un anneau, mais lui aussi d'un semi-anneau sur magma.

Le scalaire ordinaire n'est alors qu'un cas particulier du pseudo-scalaire (c'est le cas où le semi-anneau sur magma devient anneau)

Pour certains calculs, on pourra utiliser successivement une opération extérieure à pseudo-scalaire, puis une opération extérieure à scalaire normal .

3.2 - Propriétés élémentaires relatives à la loi extérieure de multiplication

Soient en lettres grecques les pseudo-scalaires pris dans ε et soient en lettres latines les éléments du semi-anneau de base, E.

Les propriétés sont les suivantes :

- a) distributivité par rapport aux pseudo-scalaires $(\alpha + \beta)x = \alpha x + \beta x$

- b) distributivité par rapport aux éléments de E

$$\alpha(x + y) = \alpha x + \alpha y$$

- c) Neutralité pour l'élément unité de ε

$$1.x = x$$

- d) annulation pour multiplication par l'élément zéro de ε ou de E

$0.x = 0$ et aussi $\alpha.0 = 0$ et aussi $0.0 = 0$

- e) associativité, dans un produit de deux facteurs, des pseudo-scalaires entre eux et des éléments de E entre eux

$$(\alpha x)(\beta y) = (\alpha\beta)(xy)$$

- f) associativité relative à un produit de scalaires

$$\alpha(\beta x) = (\alpha\beta)x$$

- g) quant à laneutralité relative à l'élément unité de E, (qu'impliquerait $\alpha I = \alpha$) nous ne l'introduisons pas comme un axiome nécessaire. Cette réserve est la seule qui introduise une dissymétrie entre les rôles de E et de ε.

3.3 - Conséquence pour le calcul d'un produit simple et d'un produit contracté

3.3.a - Produit simple

Le calcul d'un produit simple du type par exemple

$$\alpha x((\beta y\ \gamma z)\delta u)$$

autorise un regroupement des pseudo-scalaires entre eux et des éléments de E entre eux, les parenthétisations étant les mêmes de part et d'autre

$$\alpha x((\beta y\ \gamma z)\delta u) = \alpha((\beta\gamma)\delta).x((yz)u)$$

Il n'y a pas associativité du produit dans E, ni dans ε, mais on peut regrouper les termes comme ci-dessus ; et inversement, si scalaire et élément de E ont même structure, on peut distribuer le scalaire selon les éléments de E correspondants. Exemple

$$\alpha((\beta.1)\delta).x((yz)u = \alpha x((\beta y.1z)\delta u$$

3.3.b - Conséquences pour le calcul d'un produit contracté

Le produit contracté Π $(\alpha_1 a_1, \alpha_2 a_2 \ldots \alpha_n a_n)$ est égal à la somme de tous les monômes produits que l'on peut former avec $a_1 a_2 \ldots a_n$ dans cet

ordre avec toutes les parenthétisations possibles, chacun d'eux étant précédé d'un produit de $\alpha_1\alpha_2 \ldots \alpha_n$, avec la même parenthétisation.

3.4 - Un semi anneau sur magma particulier : cas d'un ensemble ε fini

Si l'ensemble ε est fini, les semi-anneaux sur magma que l'on peut définir sur ε sont isomorphes à des grammaires binaires *context free* dont les éléments non terminaux seraient des éléments de ε.

Considérons en effet toutes les équations de définition d'une telle grammaire

$$\left\{ \begin{array}{l} '''''' \\ \alpha_j \rightarrow \alpha_j\alpha_k \\ '''''' \end{array} \right.$$

Ces équations peuvent être considérées comme définissant la table de multiplication d'un semi-anneau

$$\left\{ \begin{array}{l} ''''''' \\ \alpha_j\alpha_k \rightarrow \alpha_i \\ ''''''' \end{array} \right.$$

Les couples non représentés dans cette table seront réputés avoir un produit nul.

Exemple : langage miroir, défini par $E = \{A, S_1, S\}$ et $S \rightarrow AS_1$ et $S_1 \rightarrow SA$

La table de multiplication correspondante s'écrit :

$AS_1 \rightarrow S$

$SA \rightarrow S_1$

et

$AS \rightarrow 0$

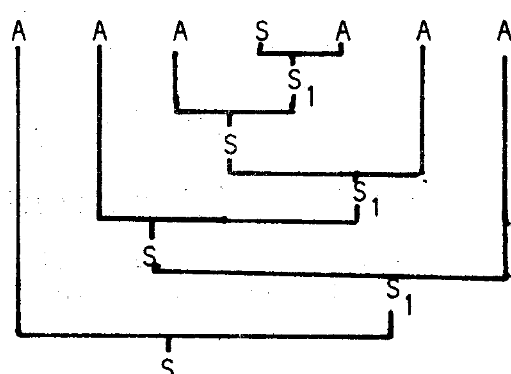
$SS_1 \rightarrow 0$

$S_1A \rightarrow 0$

$S1S \rightarrow 0$

Les seules séquences dont une factorisation produit soit égale à S sont les exemples du type : A^NSA^N

Exemple :



On notera que dans $\Pi(A^3SA^3)$ un seul terme est non nul, c'est le terme (A((A((A (S A)) A)) A)) qui correspond au graphe ci-dessus, et dont la valeur est S ; et plus généralement les seuls produits contractés égaux à S seront du type $\Pi(A^NSA^N)$, leur valeur étant fournie par un seul monome (la structure du mot miroir correspondant), et égale à S.

Supposons que nous utilisions ε comme ensemble des pseudo-scalaires pour un ensemble E = {a,c} où a aura comme coéfficient A, et c comme coefficient S. On aura par exemple

$$\Pi((Aa)^3(Sc)(Aa)^3) = S(a((a((ca)a))a))$$

et ce produit nous donne à la fois la structure retenue, et sa valeur grammaticale (Structure en lettres minuscules ; valeur grammaticale par la majuscule en facteur)

3.5 - Le produit contracté comme fonction générale de compilation

Le produit contracté apporte une formalisation de la fonction de compilation pour tout langage de structure *context free* binaire : en effet, il suffit d'affecter chaque élément terminal (mot) de la séquence (phrase) à compiler d'un speudo-scalaire égal à la somme des non terminaux susceptibles de lui correspondre ; et d'ajouter un mot supplémentaire, le "point final de la phrase", dont la catégorie sera P_t tel que :

$$SP_t \rightarrow S_f$$

et

$XP_t \rightarrow 0$ pour tout y différent de S

Alors le produit contracté sera égal à S facteur de la somme des structures autorisées par la grammaire.

Il faut signaler que cette propriété de représenter une fonction de compilation s'étend au delà des *context free* ; soit en effet une grammaire infinie du type, par exemple

nombre infini de règles {
....
$(Prod\ \alpha_i\alpha_j\alpha_k)\ (Prod\ \alpha_l\ \ldots\ a_r) \rightarrow$ (Prod chaîne fonction récursive du 1er membre)
....

mais dont chaque règle serait une phrase d'un langage dont la grammaire serait par ailleurs finie ; si une telle grammaire régissait les produits pseudo-scalaires, le calcul de compilation resterait possible; or une grammaire comme celle qui vient d'être décrite sort complètement des grammaires *context free*.

- CHAPITRE 4 -

LE "PRODUIT CONTRACTE", FORMALISATION DE LA FONCTION "EFFET PHRASES CACHEES"

Nous montrerons la chose uniquement dans le cas de grammaires context free, mais le champ d'application est plus étendu, comme indiqué plus haut.

4.1 - L'"effet phrases cachées" limité strictement aux seules phrases cachées de catégorie "S"

Ajoutons à la grammaire context free, qui est supposée donnée dans le cadre de notre exemple, un élément supplémentaire que nous dénommerons "point initial", un élément supplémentaire que nous dnommerons "point final", un élément supplémentaire X ; ajoutons également les relations suivantes :

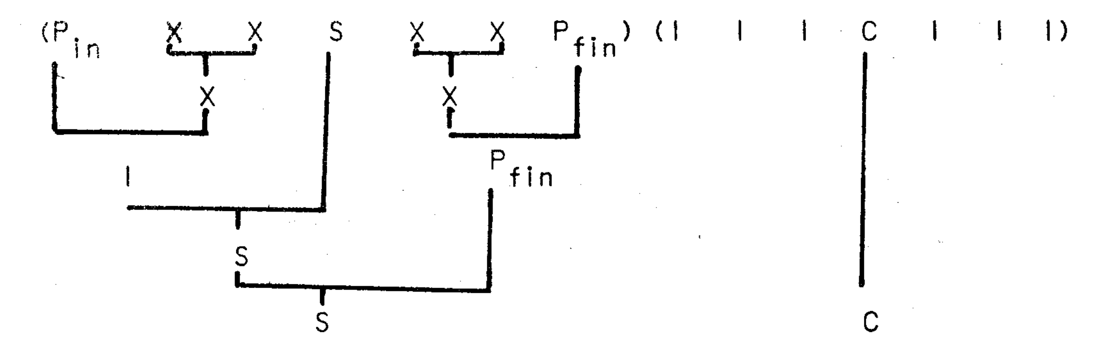
$$XX \rightarrow X$$
$$P_{in}X \rightarrow I \text{ (élément unité)}$$
$$XP_{fin} \rightarrow P_{fin}$$

Associons d'autre part, à tous les mots des phrases à analyser, le terme supplémentaire XI

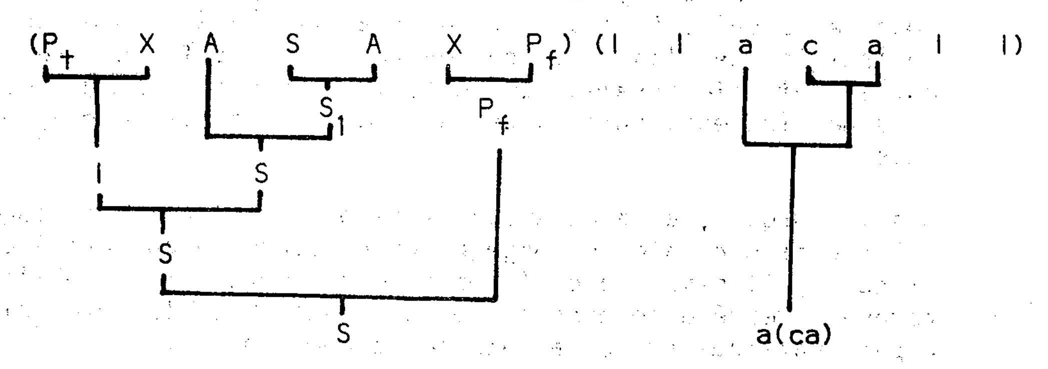
exemple :

$\Pi(a^2ca^2)$ devient ainsi $M(P_{in}(Pa + XI)^2(Sc + XI)(Aa + XI)^2\ P_{fin})$

les monomes qui subsistent seront les suivants : d'une part des monomes Sc



ou bien des monomes S(a(ca))



ou bien des monomes a^2ca^2 (schéma déjà donné)

Autrement dit, le produit contracté va nous donner (un certain nombre de fois d'ailleurs, mais il importe peu)

S(a((a(ca))a) + a(ca) + c)

qui est la somme des phrases cachées et de la phrase donnée.

4.2 - "Effet phrases cachées" étendu à d'autres groupes de mots

On peut souhaiter étendre la notion de"phrases cachées" par exemple à des chaînes nominales.

Les relations mettant en jeu la catégorie X restent les mêmes : mais les relations faisant intervenir le point final sont maintenant au nombre de deux (P_t représentant le point)

$$SP_t \rightarrow S_f \quad \text{et} \quad N_pP_t \rightarrow N_{pf}$$

alors le produit contracté donne,en plus de la structure de phrase complète, toutes les structures connexes partielles (portant sur un segment connexe partiel de phrase) susceptibles d'avoir la valeur N_p ou la valeur S.

4.3 - Caractère prévisible de "l'effet phrases cachées" selon le formalisme utilisé

Il faut bien voir que les formules ci-dessus recouvrent une réalité sous jacente plus importante que le "deus es machina" qui permet de conserver tantôt des chaînes phrases, tantôt des chaînes nominales, etc.

Cette réalité sous-jacente est celle des Π_k d'ordres $k \leq n$ dont le calcul est nécessaire pour arriver à Π_n ; avec cette contrainte que l'on doit pour évaluer Π_n calculer non seulement les Π_k qui jouent effectivement un rôle, mais encore ceux qui interviennent dans des facteurs nuls, pour vérifier justement que ces facteurs sont nuls.

C'est pourquoi, à la question de savoir si "l'effet phrases cachées" est une face cachée de langage, un discours second caché dérrière le discours, il semble que l'on puisse répondre ceci, que"l'effet phrases cachées" représente dans le langage la face de la compilation, un aspect particulier de la face de la compilation.

-=-=-=-=-=-=-=-=-=-=-=

Bulletin Technique n° 17
Institut d'Intelligence Artificielle
Université PARIS VIII

MÉTHODES DE CALCUL POUR UNE EVALUATION INFORMATIQUE DE L'EFFET "VERSIFICATION CACHÉE DANS LA VERSIFICATION"

Yves LECERF

I - OBJET DU CALCUL : LES VERSIFICATIONS CACHEES

Dans la préface de son livre "Les yeux d'Elsa", Aragon donne un commentaire sur certaines propriétés formelles singulières de plusieurs des poèmes inclus dans ce recueil, et cite notamment la sixième strophe de son poème "La nuit de Mai" :

O revenants bleus de Vimy vingt ans après
Morts à demi Je suis le chemin d'aube hélice
Qui tourne autour de l'obélisque et je me risque
Où vous errez Malendormis Malenterrés.

Comme le fait observer Aragon, cette strophe, dont il est l'auteur, et qui se compose de quatre alexandrins, peut aussi bien se réécrire sous la forme de six vers octosyllabiques :

O revenants bleus de Vimy
Vingt ans après morts à demi
Je suis le chemin d'aube hélice
Qui tourne autour de l'Obélisque
Et je me risque où vous errez
Malendormis Malenterrés.

Autrement dit, dans ce poème, une versification différente est cachée dans la versification apparente ; dans ce poème, un autre poème est caché.

D'autres exemples de propriétés formelles analogues sont, dans la même préface, donnés par Aragon, exemples tirés aussi bien de ses propres oeuvres que de celles d'autres auteurs (Racine, Boileau).

Nous nous proposons dans la présente étude de décrire des moyens de calcul récursifs susceptibles de permettre d'appréhender, d'une manière générale, les phénomènes de "versification cachée dans la versification" ; et l'objet d'un tel calcul sera aussi bien de déceler des propriétés formelles spectaculaires couvrant l'ensemble d'une strophe ou d'un poème, comme c'est le cas pour plusieurs exemples donnés par Aragon, que de détecter des phénomènes d'ampleur plus modeste, tels que par exemple la présence d'alexandrins cachés ça et là dans un texte en prose.

L'objet du présent travail n'est donc pas de donner un commentaire de fond sur la portée que l'on peut attribuer aux "versifications cachées". Sur ce point, bornons nous à mentionner : d'une part, qu'il est recommandé de se reporter à la préface des "Yeux d'Elsa" ; et d'autre part que l'existence de ces phénomènes joue pour nous un rôle dans l'appréciation aussi de problèmes théoriques qui se situent éventuellement en dehors

du domaine de la poésie (problèmes de l'art invisible ou de l'art caché en général, dans leurs relations avec l'analyse algorithmique ; problèmes du rôle syntaxique à attribuer aux effets de phrases cachées ;)

Le problème de calcul auquel nous limitons notre propos étant celui de la détection de "versifications cachées", il ne semble pas à première vue necessaire de considérer la "versification non cachée", c'est à dire la versification donnée par l'édition des textes. Autrement dit, l'objet de notre étude ne semble pas à première vue être la construction d'une fonction de "reconnaissance" des structures poétiques normales. Mais on voit bien qu'une fonction susceptible de détecter les "versifications cachées" pourrait aussi reconnaître les structures normales si par exemple on les cachait en retranscrivant le texte en forme de prose simple. Il s'agit donc d'un seul et même problème.

La fonction récursive que nous comptons utiliser pour les calculs qui vont suivre a déjà fait de notre part l'objet d'une description, à propos de son emploi pour reconnaître des structures syntaxiques : il s'agit d'un "produit contracté" défini sur une algèbre de semi-anneau sur un magma E, avec opération extérieure de produit par un "pseudo scalaire" défini sur un second semi-anneau $\mathcal{E}$. On pourra se reporter pour le détail de ses propriétés à notre étude "l'effet phrases cachées (cf. supra)", mais les définitions de cette algèbre seront succintement données à nouveau ici.

Il n'est pas indifférent que la même fonction soit proposée pour l'analyse des structures syntaxiques des phrases et pour l'analyse des structures de versification des poèmes : un tel résultat nous semble conforme à l'idée que nous nous faisons des rapports entre la poésie et le langage.

En outre, l'emploi de la même fonction de calcul présente un intérêt technique: il peut arriver en effet que certaines contraintes de versification fassent intervenir des paramètres syntaxiques (exigence, par exemple, que chaque strophe se termine par un point final de phrase).

Le déroulement de l'étude sera donc le suivant : on considérera la fonction proposée, on la confrontera successivement à une série de problèmes de reconnaissance de structures de difficultés croissantes, en montrant chaque fois que la fonction proposée convient. Les problèmes abordés seront : d'abord celui de la longueur syllabique des vers, en faisant intervenir des structures de strophes composées de vers de longueur différente, et en tenant compte des valeurs syllabiques variables de certains mots ; le problème de la scansion ne sera que très rapidement évoqué, on montrera qu'il se traite comme le précédent ; ensuite on examinera les problèmes de rimes ; enfin les problèmes de contraintes mixtes faisant intervenir à la fois la versification et les données syntaxiques.

Le fait de montrer qu'une même fonction récursive, initialement proposée pour la compilation de structures syntaxiques, soit apte aussi à effectuer ces différents calculs de structure de versification ; et que cette même fonction ait été proposée pour détecter les structures syntaxiques de "phrases cachées"; et qu'elle soit aussi proposée pour détecter les versifications cachées, voire les versifications cachées dans les phrases cachées, tout cela constitue à nos yeux un ensemble de concordances non dépourvu de portée théorique ; et cela donne à posteriori une justification supplémentaire de notre entreprise à première vue singulière, de calcul de versification cachée dans la versification.

II - FONCTION PROPOSEE POUR LE CALCUL.

II,1) Fonction proposée.

La fonction proposée pour le calcul est la fonction $\prod$,"produit contracté dans un magmoïde", c'est à dire, produit contracté dans un semi-anneau A défini sur un magma E avec opération extérieure de multiplication par un pseudo scalaire appartenant à un autre semi-anneau α défini sur un magma $\mathcal{E}$.

Sa formule de définition récursive est :

$$\begin{cases} \prod (a_1\ a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \prod (\xrightarrow{i} a_1\ a_2 \ldots a_n)\ \prod (a_1\ a_2 \ldots a_n \xleftarrow{n-i}) \\ \prod (ai) = ai \quad \text{pour tout } ai \in E \end{cases}$$

formule dans laquelle les produits et les sommes doivent s'interpréter en termes d'opérations de semi-anneau E ; dans laquelle $(\xrightarrow{i} a_1\ a_2 \ldots a_n)$ désigne la chaîne $a_1\ a_2 \ldots a_n$ limitée à ses i premiers termes; et dans laquelle $(a_1\ a_2 \ldots a_n \xleftarrow{n-i})$ désigne la chaîne $a_1\ a_2 \ldots a_n$ limitée à ses n-i derniers termes.

La structure algébrique de "magmoïde" a fait l'objet d'une définition détaillée dans l'étude "méthodes de calcul pour une analyse informatique des structures syntaxiques de textes poétiques (effet phrases cachées)" Cf. supra.

Nous rappelons toutefois quelques définitions.

II,2) Algèbre de semi-anneau sur un magma E.

On dira qu'un ensemble E est doté d'une structure de semi-anneau sur un magma si cet ensemble est doté de deux lois de composition respectivement dénommées "addition" et "multiplication", satisfaisant aux axiomes suivants :

- pour la loi de multiplication, E est un magma (pas d'associativité ni de commutativité) avec un élément neutre (élément unité 1 neutre à droite et à gauche) et un élément zéro (tous produits par zéro à gauche ou à droite égaux à zéro) ;

- pour la loi d'addition, E est monoïde commutatif avec un élément neutre zéro (qui est aussi mentionné pour la multiplication) ;

- la multiplication est distributive à gauche et à droite par rapport à l'addition.

II,3) Magmoïde.

On dira qu'un ensemble E est doté d'une structure de magmoïde avec opération extérieure de multiplication par des pseudo scalaires d'un ensemble $\mathcal{E}$, si :

- E est doté d'une structure de semi-anneau sur un magma
- $\mathcal{E}$ également
- il existe une opération extérieure de multiplication pseudo scalaire des éléments de E par des pseudo scalaires pris dans $\mathcal{E}$; satisfaisant aux axiomes suivants : distributivité par rapport aux éléments de E; distributivité par rapport aux éléments de $\mathcal{E}$;associativité dans

un produit de deux facteurs des pseudo scalaires entre eux et des éléments de E entre eux, chacun des deux produits restant en lui-même non associatif ; neutralité pour l'élément unité de $\mathcal{E}$; produit nul pour l'élément zéro de $\mathcal{E}$ ou pour celui de E.

II,4) Propriétés remarquables de la fonction "produit contracté" sur un magmoïde $\Pi(a_1\ a_2\ \ldots\ a_n)$

Le produit contracté est distributif par rapport à l'addition des pseudo scalaires comme par rapport à celle de E.

Il est récursif, sous réserve que les sommes et produits dans E et $\mathcal{E}$ soient récursifs.

Il est canoniquement égal à une somme de termes, couvrant l'ensemble de toutes les parenthétisations produits de la chaîne $a_1\ a_2\ \ldots\ a_n$; (dans chacun de ces termes, on peut mettre l'ensemble des scalaires en facteur avec même parenthétisation) sauf lacunes pour les termes à coefficients scalaires nuls.

Si $\mathcal{E}$ est fixé, on peut lui associer une famille de langages context free dont Π est fonction de compilation ; Π compile également les intersections de ces langages.

Si $\mathcal{E}$ est infini, les opérations produit et somme sur E restant cependant récursives, on peut lui associer une famille de langages à grammaire "context free infinie" ; Π compile ces langages et leurs intersections.

La particularité de ces compilations par l'intermédiaire de Π , consiste dans le fait que le résultat se présente sous forme d'une somme de monômes qui chacun représentent : par les éléments de E, des structures, et, par les pseudo scalaires, les "valeurs" grammaticales de ces structures ; étant entendu que seuls les termes à "valeurs" interessant l'objet du calcul sont sélectionnés.

Dans une recherche de segments de versification cachés, le résultat se présentera comme une somme de termes monomes constitués par les chaînes retenues, avec des coefficients pseudo scalaires indiquant la "valeur" de versification de chacun de ces segments.

II,5) Rappel des motifs pour lesquels cette fonction est proposée.

Le motif, est que cette fonction a déjà été proposée par ailleurs comme fonction représentative de la compilation d'analyse syntaxique des langues naturelles (voir article déjà cité plus haut en référence). Pour des raisons théoriques et aussi pour permettre l'expression de contraintes mixtes portant à la fois sur la versification et la syntaxe, on souhaite pouvoir utiliser la même fonction.

III - PRISE EN COMPTE DES FORMES LES PLUS ELEMENTAIRES DE CONTRAINTES DES LONGUEURS SYLLABIQUES DES VERS ET DES CONTRAINTES IMPOSANT DES TYPES DE SCANSION DETERMINES.

III,1) Description des contraintes à prendre en compte en matière de longueurs syllabiques.

Considérons le poème de Valéry "Ni vu ni connu" :

Ni vu ni connu
Je suis le parfum
Vivant et défunt
Dans le vent venu !

et cherchons un mécanisme susceptible de rendre compte de la divisibilité de ce texte en segments-vers pentasyllabiques, le texte étant supposé donné à priori en prose.

Si nous associons aux différents mots des coefficients représentant le nombre de syllabes qu'ils comportent,

Ni = 1
Vu = 1
Connu = 2
Je = 1
Suis = 1
Le = 1
Parfum = 2
etc...

la condition à remplir est la suivante : pouvoir placer dans le texte des interrupteurs ⊗ de telle manière que les distances les séparant soient égales à cinq syllabes :

⊗	Ni	vu	ni	connu	⊗	Je	suis	le	parfum	⊗	Vivant	et	défunt	⊗	Dans	le	vent	venu
	1	1	1	2		1	1	1	2		2	1	2		1	1	1	2

Formulée dans ce mode simpliste, la contrainte syllabique relève d'un simple mécanisme d'automate fini. Pour des raisons qui seront précisées plus loin, nous utiliserons toutefois, même pour cette contrainte très simple, un formalisme de grammaire générative de C-langage ; ce qui est possible, les langages de Kleene étant des C-langages particuliers.

Toujours dans le cadre des études de contraintes syllabiques, nous devrons faire mention des problèmes de prise en compte des syllabes à e muet.

Pour en donner un exemple, considérons deux alexandrins de Racine : soient les vers 457 et 458 de la tragédie Andromaque :

"Je n'ai point du silence affecté le mystère
Je croyais sans péril pouvoir être sincère"

et associons à chaque mot un coefficient indiquant le nombre de syllabes qui le composent :

"Je	n'ai	point	du	silence	affecté	le	mystère
1 0	1	1	1	2	3	1	2

Je croyais sans péril pouvoir „être sincère"
1 2 1 2 2 2 2

La condition pour que ce texte soit en alexandrins peut s'écrire :

$$\sum \text{coeff} = 12$$

ce que l'on vérifie en effet : $1 + 0 + 1 + 1 + 1 + 2 + 3 + 1 + 2 = 12$

$1 + 2 + 1 + 2 + 2 + 2 + 2 = 12$

mais en observant que l'on a compté :

silence = 2 (mot suivi d'une voyelle)
être = 2 (mot suivi d'une consonne)
mystère = 2 (mot en finale de vers)
etc...

III,2) Calculs préliminaires sur l'arithmétique des longueurs de chaînes.

soient les relations :

$C_1 \longrightarrow m_1$
$C_2 \longrightarrow C_1\ C_1$ ou m_2
$C_3 \longrightarrow C_1\ C_2$ ou $C_2\ C_1$ ou m_3
$C_4 \longrightarrow C_1\ C_3$ ou $C_2\ C_2$ ou $C_3\ C_1$ ou m_4
etc...

interprétables comme suit :

$C_1 \longrightarrow m_1$ signifie : on peut constituer une chaîne de une syllabe seulement par un mot de une syllabe.

$C_2 \longrightarrow C_1$ ou m_2
signifie : on peut constituer une chaîne de deux syllabes avec : deux chaînes de une syllabe, ou un mot de deux syllabes.

$C_3 \longrightarrow C_2\ C_1$ ou $C_1\ C_2$ ou m_3
signifie : on peut constituer une chaîne de trois syllabes avec : une chaîne de deux syllabes et une chaîne de une syllabe, ou bien une chaîne de une syllabe et une chaîne de deux syllabes, ou bien un mot de trois syllabes.

etc, les C_i désignant des chaînes de mots, longues au total de i syllabes, et les m_j désignant des mots simples longs de j syllabes.

Cette série de relations représente tout simplement une table d'addition des longueurs syllabiques, dans la forme toutefois remarquable d'une grammaire context free.

III,3) Découpage d'un texte en chaînes de même syllabicité.

Ce problème de découpage en chaînes est une première approche de celui du découpage en vers ($\neq$ de chaîne) de même syllabicité. Nous écrirons, pour une syllabicité 5, par exemple :

S (texte) $\longrightarrow C_5$ S S $\longrightarrow$ point final	condition de découpage en chaînes de cinq syllabes	Problème : ces conditions sont-elles satisfaites par le texte S donné ?
$C_1 \rightarrow m_1$ $C_2 \rightarrow C_1\ C_1$ ou m_2 etc............ $C_5 \rightarrow C_1\ C_4$ ou...	grammaire des chaînes (utile jusqu'à C_5)	

Ces relations signifient : pour les deux premières que le texte se compose d'une série de C_5 jusqu'au point final ; le découpage se ramène à une compilation context free.

Or, nous savons que la fonction Π proposée permet de compiler les grammaires context free.

Cette méthode permettrait par exemple le découpage en vers du poème de Valéry

Ni vu ni connu
Je suis le parfum
Vivant et défunt
Dans le vent venu

Ni vu ni connu
Hasard ou génie
A peine venu
etc...

en faisant :

Ni, vu, je, suis, le, et, dans, vent, ou, à, etc... $\longrightarrow m_1$

Connu, parfum, vivant, défunt, venu, hasard, génie, peine, $\longrightarrow m_2$

III,4) <u>Prise en compte des finales de mots en e muet.</u>

Un mot M se compose d'un début δ (voyelle δa ou consomme δb), d'un corps de mot m, et d'une fin de mot φ (muette φb ou néant φa). On peut écrire :

$$\frac{M}{\delta\, m_j\, \varphi}\quad \frac{M}{\delta\, m_r\, \varphi}$$

$\varphi a\ \delta a$ = mo (o syllabe)
$\varphi a\ \delta b$ = mo (o syllabe)
$\varphi b\ \delta a$ = mo (o syllabe)
$\varphi b\ \delta b$ = m1 (1 syllabe)

Il suffit de réécrire des équations du type :

$V_1 \longrightarrow \delta C_1 \varphi$ ou $\delta m_1 \varphi$
$V_2 \longrightarrow \delta C_2 \varphi$ ou $\delta C_1\ C_1 \varphi$ ou $\delta m_2 \varphi$ (V_i signifiant vers de i pieds)
$V_3 \longrightarrow$ etc...

en donnant aux δ et φ les deux valeurs possibles chaque fois, pour que cette nouvelle grammaire suffise à prendre en compte le problème des finales muettes

III,5) Découpage d'un texte en vers de syllabicité variable.

Considérons le poème de Musset :

Dans Venise la rouge
Pas un bateau qui bouge
Pas un pêcheur dans l'eau
Pas un falot

Sa composition, notée en terme de vers V_i à longueur corrigée pour tenir compte des e muets, s'écrit :

$$G \begin{cases} S \longrightarrow V_6\ V_6\ V_6\ V_4\ S \\ S \longrightarrow \text{point final} \end{cases}$$

qui donne : $S \longrightarrow V_6\ V_6\ V_6\ V_4 \quad V_6\ V_6\ V_6\ V_4 \quad V_6\ V_6\ V_6\ V_4$ etc..$V_6\ V_6\ V_6\ V_4$

G n'est pas vraiment une grammaire context free binaire, mais se ramène facilement à une grammaire binaire. Il faut considérer que les relations de longueur définissant les V_i font aussi partie de la grammaire.

Considérons comme autre exemple le poème de Victor Hugo "Les Djinns"

1 Murs, ville
Et port
Asile
......
(8 vers)

2 Dans la plaine
Nait un bruit
C'est l'haleine
....
(8 vers)

3 La voix plus haute
Semble un grelot
D'un nain qui saute
....
(8 vers)

On sait que la composition des strophes passe successivement par des longueurs de vers de 2, 3, 4, 5, 6, 7, 8 et 10 syllabes, pour 8, 7, 6, 5, 4, 3, 2, en redescendant. La composition du poème peut s'écrire :

$$S \longrightarrow V_2^8\ V_3^8\ V_4^8\ V_5^8\ V_6^8\ V_7^8\ V_8^8\ V_{10}^8\ V_8^8\ V_7^8\ V_6^8\ V_5^8\ V_4^8\ V_3^8\ V_2^8$$

Equation context free qui n'est pas binaire mais peut se ramener à une série de relations binaires.

III,6) Prise en compte des contraintes imposant des types de scansion déterminés.

Ces contraintes ne sont pas toujours explicitées, ou même pas toujours prises en considération. Elles font intervenir l'accent tonique, les longues et les brèves. Elles ont toujours un rôle explicitement important pour les versifications grecques et latines.

Dans la mesure où les types sont déterminés, et en nombre fini, les mécanismes sont de même nature que ceux relatifs par exemple à la composition des strophes constituant un poème, et il ne semble pas necessaire de nous étendre plus longuement sur ce point.

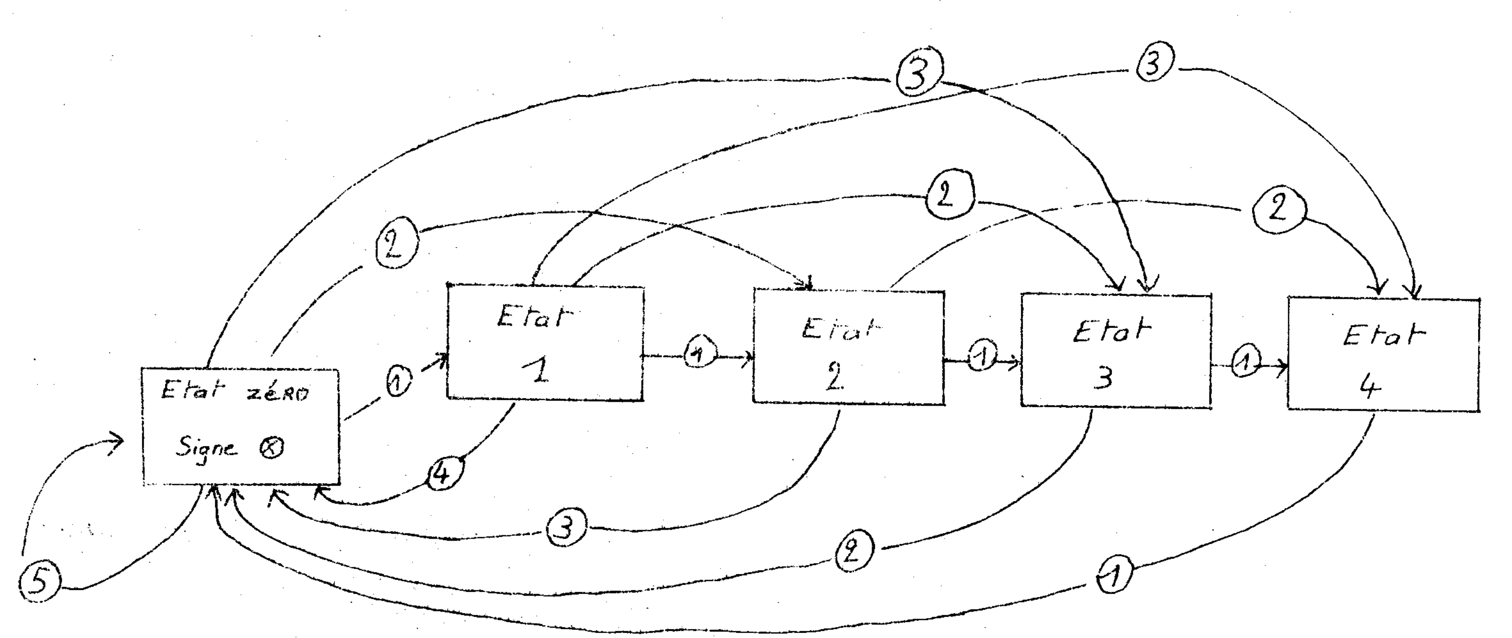
Si on détecte dans un texte un vers isolé, il faut au moins qu'il ait des propriétés remarquables de scansion pour mériter d'être signalé.

III,7) Discussion sur le choix d'une représentation par C-langage plutôt que par automate fini.

L'ensemble des mécanismes de prise en compte des longueurs syllabiques de vers qui viennent d'être décrits relève visiblement de l'ordre des mécanismes d'automates finis.

Nous avons choisi la formalisation context free pour la raison qu'elle donne une formulation beaucoup plus concise.

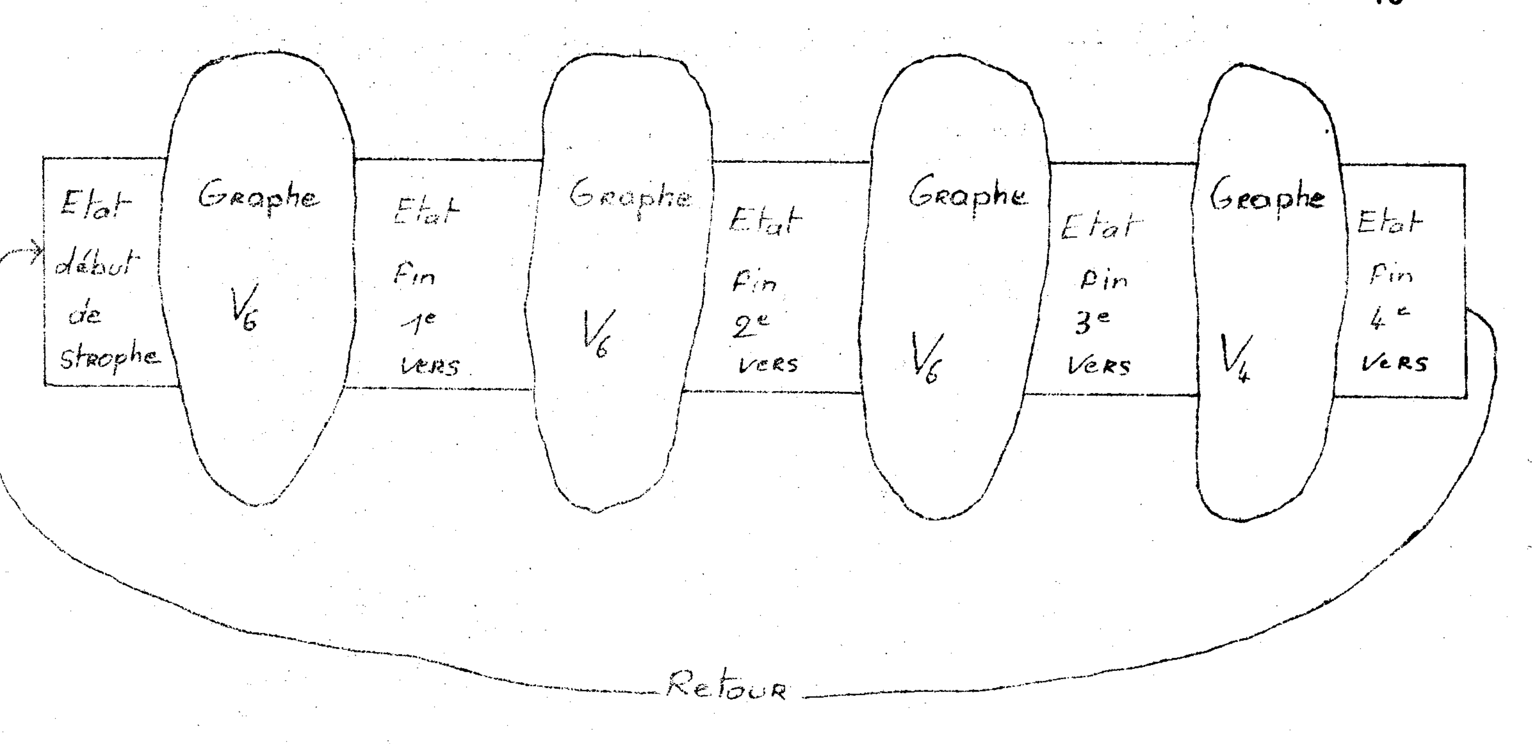
Ci-après, un graphe indiquant la complexité du problème de découpage en chaînes de cinq syllabes (sans correctifs de e muet) traité en automate fini.



Appelons ce graphe du nom de "Graphe V_5", puisqu'il représente un découpage V_5 ; la structure d'un "Graphe V_6" ou d'un "Graphe V_4" est très analogue et se déduit facilement de celle du "Graphe V_5" ; on peut ensuite combiner des graphes de ce genre pour représenter des successions de vers de longueurs différentes ; le graphe représentatif d'un découpage $V_6V_6V_6V_4$ comme dans le poème de Musset

Dans Venise la rouge
Pas un bateau qui bouge
Pas un pêcheur dans l'eau
Pas un falot

serait environ quatre fois plus volumineux, et nous ne pouvons donner ci-après que sa structure globale articulée en trois "Graphes V_6" et un "Graphe V_4"



IV - PRISE EN COMPTE DES FORMES LES PLUS ELEMENTAIRES DE CONTRAINTES EXPRIMANT UN PHENOMENE D'ECHO DU TEXTE DANS LUI-MEME (CONTRAINTES DE RIME ; CONTRAINTES DE VERSIFICATION).

IV,1) Généralités sur le langage écho.

Considérons des phrases représentant le dialogue que l'on peut avoir avec un écho :

Bonjour / Bonjour (écho)
Comment vas-tu / Comment vas-tu (écho)
etc...

le langage constitué par l'ensemble des phrases de ce genre, sur un alphabet terminal A, peut se représenter par les formules :

$$\begin{cases} \forall x \in A^{*} \quad xx \in L \\ L = \{xx\} \quad \text{avec } x \in A^{*} \end{cases}$$

C'est ainsi que nous pouvons doubler en écho le poème de Valéry déjà cité

Ni vu ni connu
Ni vu ni connu
Je suis le parfum
Je suis le parfum
Vivant et défunt
Vivant et défunt
Dans le vent venu
etc...

Ce résultat n'a rien d'aussi anormal que l'on pourrait imaginer : bien des formes poétiques et beaucoup de chansons populaires, qui sont d'une certaine façon des poèmes, contiennent des parties de phrases que l'on répète. Par exemple :

Aux marches du Palais
Aux marches du Palais
Y a un' tan belle fille
Lon la
Y a un' tan belle fille

Elle a tant d'amoureux
Elle a tant d'amoureux
Qu'elle ne sait lequel prendre
Lon la
Qu'elle ne sait lequel prendre

dans la chanson du Quadrille des Homards (Lewis Carrol, "Alice au pays des merveilles"), on trouve des vers tels que :

Voulez-vous, ne voulez-vous pas, voulez-vous, ne voulez-vous pas
(rejoindre la danse etc...)

et plus loin :

Il ne veut pas, il ne peut pas, il ne veut, il ne peut pas
(rejoindre la danse etc...)

ou bien dans un virelai d'Eustache Deschamps :

Suis-je, suis-je, suis-je belle etc...

Du point de vue de la théorie des langages, c'est un résultat connu que le langage écho ne peut pas se formaliser avec un context free.

Par contre, dans le formalisme de calcul que nous avons proposé, ce langage peut être pris en compte.

IV,2) <u>Diversité des structures ou contraintes qui peuvent relever d'une analyse en langage écho.</u>

a) Structures de rimes : écho partiel sur la fin des vers.
L'identité des fins de séquences correspond à une structure du type :

AX, BX

Le problème de la prise en compte de la rime fera l'objet d'une discussion un peu plus loin dans cet article.

b) Structures de "rime par le début" : écho sur le début du vers.
Avec les notations précédentes, cette structure s'écrit :

XA, XB

Sans doute l'idée commune que l'on se fait de la rime donne le sentiment que la "rime" de vers par leur début et non par leur fin n'est pas dans les usages poétiques normaux. Pourtant, les exemples en sont innombrables

O mon âme, le soir est triste sur hier
O mon âme, le soir est morne sur demain
O mon âme, le soir est grave sur toi-même

Henri de Régnier, "Tel qu'en songe"

ou bien chez Christine de Pisan, Ballade XI :

Seulette suis et seulette veux être
Seulette m'a mon doux ami laissée
Seulette suis sans compagnon ni maître
Seulette suis dolente et courroucée etc...

et tous les 25 vers ainsi, sauf le début de l'envoi qui commence classiquement par "Prince".

c) Structure de répétition d'un vers entier, toujours le même, en finale de strophes successives, qui constitue pour la strophe un équivalent de ce que la rime est pour le vers.

C'est par exemple ce que l'on trouve dans les Ballades : Ballade de François Ier, dont les strophes se terminent par "quand de te voir j'ai perdu l'espérance" ; Ballade de Marot "de s'amie bien belle", dont chaque strophe se termine par "c'est bien la plus belle de France".

$Strophe_A$ X , $Strophe_B$ X , $Strophe_C$ X , $Envoi_D$ X

d) Structure de répétition d'un vers entier, toujours le même, en début de strophe. Exemple : "Il est un port" (Henri de Régnier dans "Tel qu'un songe" début de deux strophes successives. Ou bien une chanson anonyme du XVIII° siècle :

Oh j'ai vu, j'ai vu
Compèr' qu'as-tu vu
etc...

} Les deux mêmes vers répétés au début des quatre strophes de la chanson (strophe de six vers)

ou bien chez Musset :

"Nous l'avons eu, votre Rhin allemand..."
(vers mis au début de quatre strophes successives).

e) Structure de répétition en début de chaque strophe non pas d'un vers entier, mais d'un début de vers, toujours le même. Par exemple chez "Bonaventure des Periers"

Si Amour n'était volage / etc...
Si Amour avait connaissance / etc...
Si Amour découvrait sa vue / etc...
Si Amour ne portait les flèches / etc...

ou bien "Revue romantique" de Musset, ou huit strophes successives commencent par le mot "heureux".

f) Répétition d'un vers en première et troisième position d'un quatrain :

Le vôtre est rendu
Je n'en ai plus d'autre
Le vôtre est rendu
Le mien est perdu
Marcelline Desbordes-Valmore.

cette structure peut se trouver à un point quelconque de la strophe :

Voulez-vous, ne voulez-vous pas, voulez-vous, ne voulez-vous pas
Voulez-vous vous joindre à la danse
Voulez-vous, ne voulez-vouspas, voulez-vous, ne voulez-vous pas
Ne voulez-vous pas vous joindre à la danse
Lewis Carroll.

ou bien dans la chanson anonyme :

Les juges de Grenoble
Avec leurs longues robes
Et leurs bonnets carrés, m'eurent bientôt
Vous m'entendez
Et leurs bonnets carrés, m'eurent bientôt
Jugé

ou bien chez Musset :

Mais de vous en souvenir
Prendrez-vous la peine
Mais de vous en souvenir
Et d'y revenir ?

IV,3) <u>Phénomènes d'échos faisant intervenir des paramètres constituant des abstractions du texte (paramètres de nature grammaticale, de structure, de versification, de sémantique).</u>

a) Paramètre de nature grammaticale des mots
reprenons l'exemple déjà cité de Henri de Régnier

O mon âme, le soir est triste sur hier
O mon âme, le soir est morne sur demain
O mon âme, le soir est grave sur toi-même

si nous transcrivons en termes paramétriques, nous obtenons :

O mon âme, le soir est Adj sur x

cette structure s'adapte en effet aux trois vers, en faisant successivement :
Adj : triste, morne, grave
X : hier, demain, toi-même
et l'on voit que le phénomène d'écho s'étend plus loin que le début du vers.
Ou bien encore dans le poème de Musset "Sur une morte"

Elle aurait souri, si la fleur... (strophe 5)
Elle aurait pleuré, si sa main... (strophe 6)
Elle aurait aimé, si l'orgueil... (strophe 7)

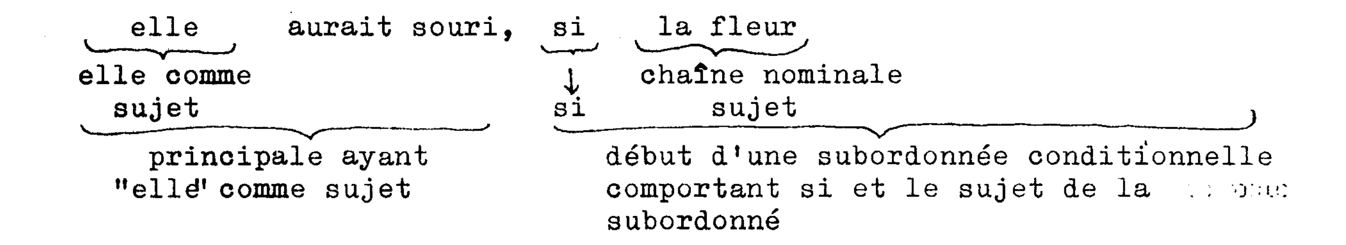
si nous transcrivons en termes paramétriques, nous observons :

Elle aurait (verbe participe passé), si (déterminatif) (nom)

cette structure s'adapte en effet aux trois exemples en faisant successivement :
verbe participe passé : souri, pleuré, aimé
déterminatif : la, sa, l'
nom : fleur, main, orgueil.

et nous avons un autre exemple d'interférence entre les catégories grammaticales et la structure de versification.

b) Paramètres de structure.
Pour le même poème, d'ailleurs, et les mêmes vers, nous aurions pu transcrire :



et cette retranscription qui est un véritable arbre de structure, s'applique alors aux six premiers débuts de strophes du poème :

Elle était belle, si la nuit etc...
Elle pensait, si le vain bruit etc...

c) Paramètre de versification.
Ce paramètre est déjà intervenu sans qu'on souligne particulièrement la chose, comme mesure naturelle du segment susceptible d'écho.

Mais c'est la notion nouvelle de "longueur égale des vers" qu'on peut vouloir considérer, non plus comme une des douze variétés de poèmes, mais comme une modulation du phénomène d'écho ; en effet, les compositions des strophes sont relativement quelconques, et la règle d'identité de versification des strophes peut s'interpréter comme une structure XX dans la versification.

d) Paramètres sémantiques.
Comme exemple d'écho sémantique, citons un étrange texte de Samuel Beckett (Watt, ed. minuit, page 47).

"Et cette pauvre pouilleuse de vieille terre, la mienne et celle de mon père et de ma mère et du père de mon père et de la mère de ma mère et de la mère de mon père et du père de ma mère et du père de la mère de mon père et de la mère du père de ma mère et de la mère de la mère de mon père et du père du père de ma mère et de la mère du père de mon père et du père de la mère de ma mère et du père du père de mon père et de la mère de la mère de ma mère et des pères et mères d'autres infortunés et des pères de leurs pères et des mères de leurs mères et des mères de leurs pères et des pères de leurs mères et des pères des mères de leurs pères et des mères des pères de leurs mères et des mères des mères de leurs pères et des pères des pères de leurs mères et des mères des pères de leurs pères et des pères des mères de leurs mères et des pères des pères de leurs pères et des mères des mères de leurs mères".

Ce texte n'est pas à proprement parler doté d'une structure de versification ; mais on peut certainement dire qu'il constitue une sorte de "litanie", batie selon une structure régulière sur la combinatoire de quelques paramètres sémantiques élémentaires.

IV,4) Discussion détaillée du problème de la rime considérée soit comme "écho partiel" sur la fin des vers soit comme une propriété d'un "langage régulier" généré par un automate fini.

Par sa nature, souvent, la rime rappelle un écho

La corde nue et maigre
Grelottant sous le froid
Beffroi
Criait d'une voix aigre
Qu'on oublie au couvent
L'Avent
Musset "Un rêve".

et d'autre part, le niveau de généralité du phénomène XX est suffisamment large pour constituer sans hésitation l'explication normale et la justification normale de la rime.

Toutefois, nous devons tout de même mentionner qu'une interprétation concurrente est possible : justification de la rime par une explication de type "context free" ou "automate fini", en exploitant le fait que le nombre de catégories différentes de rimes peut faire l'objet d'un inventaire fini.

On retombe alors dans un inventaire context free, un vers se caractérisant par son nombre de syllabes n et sa rime x s'écrit $V_n(x)$ et on peut noter un sonnet par exemple :

$$\text{Sonnet}_8 = V_8(x),\ V_8(y),\ V_8(y),\ V_8(x)\ ;\ V_8(x),\ V_8(y),\ V_8(x),\ V_8(y)\ ;$$
$$V_8(u),\ V_8(u),\ V_8(w)\ ;\ V_8(v),\ V_8(v),\ V_8(w).$$

IV,5) Intégration de la fonction "rime" dans le formalisme.

Prenons l'exemple du problème de la formalisation du sonnet, la série de formules suivantes représente un sonnet en vers de huit syllabes.

Formules	Observations
$\text{Sonnet} = \{Q\}_8\ (x,y)\quad \{T\}_8\ (u,v,w)$	
$\{Q\}_8\ (x,y) = Q_8\ (x,y)\quad Q'_8\ (x,y)$	
$\{T\}_8\ (u,v,w) = T_8\ (u,w)\quad T'_8\ (v,w)$	
$Q_8\ (x,y) = D_8\ (x,y)\quad D_8\ (y,x)$	
$Q'_8\ (x,y) = D_8\ (x,y)\quad D_8\ (x,y)$	
$D_8\ (x,y) = V_8\ (x)\quad V_8\ (4)$	règle unique transcrite une fois en xy et une fois en yx
$D_8\ (y,x) = V_8\ (4)\quad V_8\ (x)$	
$T_8\ (u,w) = D_8\ (u,u)\quad V_8\ (w)$	(pour tous u,w)
$T'_8\ (v,w) = D_8\ (v,v)\quad V_8\ (w)$	(pour tous v,w)

Des formules analogues, mais transposées sur des longueurs de vers différentes, s'en déduisent facilement pour des sonnets constitués de vers non octosyllabiques; nous donnons ci-après la transposition "sonnet" (xyuvw) correspondant à un sonnet de vers tétrasyllabiques faisant intervenir sur ses quatrains les rimes xy et sur ses tercets les rimes u,v,w ;

	Formules							Conditions ou observations
1)	$Sonnet_4$ (x,y,u,v,w)		=	$\{Q\}_4$	(x,y)	*	$\{T\}_4$ (u,v,w)	∀ (x,y,u,v,w)
2)	$\{Q\}_4$	(x,y)	=	Q_4	(x,y)	*	Q'_4 (x,y)	∀ x,y
3)	$\{T\}_4$	(u,v,w)	=	T_4	(u,w)	*	T_4 (v,w)	∀ u,v,w
4)	Q_4	(x,y)	=	D_4	(xy)	*	D_4 (yx)	∀ xy
5)	Q'_4	(xy)	=	D_4	(xy)	*	D_4 (xy)	∀ xy
6)	T_4	(ij)	=	D_4	(ii)	*	V_4 (j)	∀ ij
7)	D_4	(p,q)	=	V_4	(p)	*	V_4 (q)	∀ pq

- La première règle exprime qu'un sonnet en vers tétrasyllabiques se compose d'un groupe $\{Q\}_4$ (x,y) de quatrains de rimes x,y et d'un groupe $\{T\}_4$ (u,v,w) de tercets faisant intervenir les rimes u,v,w.

- La seconde règle exprime que le groupe $\{Q\}_4$ (x,y) de quatrains se compose d'un quatrain Q_4 (x,y) type Q (c'est à dire à rimes x,y,y,x) faisant intervenir les rimes x et y, et d'un quatrain Q'_4(x,y) de type Q' (rimes croisées x,y,x,y) faisant intervenir les mêmes rimes x et y.

- La troisième règle exprime que le groupe $\{T\}_4$ (u,v,w) de tercets se compose de deux tercets standards T_4 (u,w) et T_4 (v,w) faisant intervenir respective ment les rimes uw et vw.

- La quatrième règle exprime qu'un quatrain du type Q_4 (x,y) se compose de deux "doublets" D_4 (xy) et D_4 (yx), pour tout y et pour tout x.

- La cinquième règle exprime qu'un quatrain du type Q'_4 (xy) se compose de deux "doublets" D_4 (xy) et D_4 (xy) faisant intervenir ces mêmes rimes x et y.

- La sixième règle exprime qu'un tercet standard T_4 (ij) se compose d'un doublet D_4 (ii) et d'un vers V_4 (j) de rime j. Cette règle s'appliquera deux fois, puisqu'il y a deux tercets.

T_4 (u,w) D_4 (u,u) + V_4 (w)

T_4 (v,w) D_4 (v,v) + V_4 (w)

- La septième règle exprime qu'un doublet D_4 (p,q) se compose de deux vers V_4 (p) et V_4 (q).

La description ainsi donnée du sonnet est correcte ; on peut certainement se demander si une analyse du phénomène de rime par inventaire fini des rimes possibles est une analyse qui atteint la vraie nature du problème constituant son objet ; mais la description resterait la même pour un alphabet de rimes "infini".

IV,6) Modalités effectives du calcul.

Rappelons que la structure que nous utilisons est celle d'une algèbre de semi-anneau sur un magma E, avec multiplication extérieure par des pseudo scalaires éléments d'un autre semi-anneau $\mathcal{E}$.

a) Langage écho, dans sa forme élémentaire.

L'ensemble $\mathcal{E}$ des scalaires est constitué par des produits cartésiens d'éléments αi du monoïde libre E* , et par deux éléments : + et - . Les règles de produit sont les suivantes :

$$\alpha i > \alpha j \qquad (\sigma -, \alpha i) \otimes (\sigma -, \alpha j) \longrightarrow (\sigma -, \alpha i \alpha j)$$
$$\alpha i > \alpha j \qquad (\sigma -, \alpha i) \otimes (\sigma -, \alpha j) \longrightarrow (\sigma +, \alpha i \alpha j)$$

Ces règles de calcul sont récursives.
L'expression de l'analyse de la structure, dans sa forme finale, tire en outre parti du fait que :

$$\prod \left[\alpha i \; Xj \otimes \alpha j \; Xj\right] \longrightarrow \alpha i \alpha j \quad \prod (Xi \; Xj)$$
tant que $\alpha i \alpha j$ est en $\sigma -$

$$\prod (\alpha i \; Xi \otimes \alpha j \; Xj) = [\alpha i \alpha j] \prod (Xi) \quad \prod (Xj)$$
quand $\alpha i \alpha j$ est en $\sigma +$

b) Calcul sur un écho partiel sur demi-chaînes.

Le calcul suit le même principe : mais le scalaire associé à une chaîne i image de E ; un élément de longueur de chaîne, qui est un nombre ; un élément de fonction de chaîne qui est $\sigma -$ ou σ_o ou $\sigma +$.
La grammaire pour les éléments de fonction S est comparable à celle déjà vue plus haut .

$$\begin{array}{ll} \forall \lambda i \, \lambda j & \left\{ \begin{array}{l} (\sigma -, \lambda i, \alpha i) \otimes (\sigma -, \lambda j, \alpha j) \longrightarrow (\sigma -, \lambda i + \lambda j, \alpha i \alpha j) \\ (\sigma -, \lambda i, \alpha i) \otimes (\sigma -, \lambda j, \alpha j) \longrightarrow (\sigma_o, \lambda i, \alpha i) \\ (\sigma_o, \lambda i, \alpha i) \otimes (\sigma_o, \lambda k, \alpha k) \longrightarrow (\sigma +, \lambda i, \alpha i) \end{array} \right. \\ \text{Si } \lambda i = \lambda j & \\ \text{Si } \lambda i > \lambda k & \end{array}$$

On notera que certains produits ont deux valeurs possibles.
Meilleure solution :

le scalaire se compose du produit cartésien

- d'un élément de fonction de chaîne
- d'un élément de nature de première moitié de chaîne
- d'un élément de nature de chaîne
- d'un élément de parité de chaîne.

Les fonctions de chaînes sont en $\sigma -$ et $\sigma +$

$\sigma - \sigma - \longrightarrow \sigma -$ en général
$\sigma - \sigma - \longrightarrow \sigma +$ seulement s'il y a { identité des premières moitiés de chaîne / identité des parités }

Les parités indiquent seulement la parité des nombres de mailles de chaîne.

Les "natures de chaîne" sont définies comme précedemment : ce sont les éléments d'un monoïde image de E* .
Les natures de demi-chaîne prennent en compte la première moitié de chaque chaîne si la chaîne est paire, de longueur 2p ; si la chaîne est impaire 2p + 1 étant sa longueur, la nature de demi-chaîne prend en compte les p + 1 premiers symboles.

V - PRISE EN COMPTE DE CONTRAINTES EXPRIMANT DES COMBINAISONS DE CRITERE GRAMMATICAUX SYNTAXIQUES AVEC DES CRITERES DE VERSIFICATION.

V,1) Recensement de contraintes.

a) Début et fin de phrase et charnière de vers.

Il est au moins une contrainte évidente à l'interface su "syntaxique" et du "versifié" : le "début de texte" est à la fois "début de vers" et "début de phrase" : la "fin du texte" est à la fois "fin de vers" et "fin de phrase".

Mais les contraintes relatives aux débuts et fins de phrases vont en général bien au-delà de cette simple observation.

Pour la plupart des morceaux courts du type sonnet, ballade, ode, etc, on peut remarquer que leurs auteurs se sont imposé une contrainte de nombre entier de phrases par strophe : autrement dit le début de strophe est toujours début de phrase , la fin de strophe est toujours fin de phrase. Certains auteurs il est vrai, se donnent licence de déroger à cette règle ; mais ces dérogations obéissent alors elles-mêmes à des règles précises : par exemple, une phrase peut couvrir plus d'une strophe, mais alors elle couvrira un nombre entier de strophes et son point final coïncidera avec une fin de strophe.

Pour des morceaux longs et sans strophe , tels que les pièces de théâtre en vers, on constate aussi que des contraintes lient la versification et le découpage en phrases. Dans Racine, par exemple, beaucoup de vers sont des phrases entières :

> "C'est lui que je prétends honorer aujourd'hui"(Esther)
> "Ils mettront ma vengeance au rang des parricides" (Britannicus)
> "Je croyais sans péril pouvoir être sincère" (Andromaque)
> "Arsace, il faut partir quand j'aurai vu la reine" (Bérénice)

souvent aussi, un vers se compose de deux propositions indépendantes :

> "J'ai des yeux ; leur bonheur n'est pas encore tranquille :
> On trompe Iphigénie ; on se cache d'Achille ;
> Agamemnon gémit. Ne désespérons point ;" (Iphigénie)

Il peut arriver qu'une phrase ou proposition indépendante plus courte qu'un vers ne soit pas accompagnée d'une autre phrase courte pour compléter ce vers:

> "Oui, Seigneur, nous partions ; et mon juste courroux
> Laissait bientôt Achille et le camp loin de nous." (Iphigénie)

mais on voit que dans ce cas il subsiste pour chaque phrase la contrainte de faire coïncider son début et sa fin soit avec une charnière de versification soit avec un milieu d'alexandrin.

b) Débuts et fins de syntagmes et charnières de vers.

A défaut de coïncider avec un début ou une fin de phrase ou de proposition indépendante, la charnière de vers va généralement coïncider avec une charnière syntaxique importante de la phrase. Charnière entre principale et subordonnée relative :

"Vous avez de ses feux, un gage solennel
Qu'il peut, quand il voudra, confirmer à l'autel" (Mithridate)

charnière entre sujet et prédicat de la principale :

"Arbate, à cet hymen chargé de vous conduire,
De votre obéissance aura soin de m'instruire" (Iphigénie)

"Que faites-vous Madame ? Et quel ressouvenir
Tout à coup vous arrête et vous fait revenir ?" (Mithridate)

charnière entre sujet et prédicat d'une subordonnée :

"Hélas, ignorez-vous quelles sévères lois
Aux timides mortels cachent ici les rois ?" (Esther)

charnière d'une coordination par "et" :

"Croyez-moi, chère Esther, ce sceptre, cet empire
Et ces profonds respects que la terreur inspire
A leur pompeux éclat mêlent peu de douceur" (Esther)

"Aricie, à la fin, de son sort est maîtresse
Et bientôt à ses pieds verra toute la Grèce" (Phèdre)

Il serait difficile de formaliser sur un plan général des contraintes qui varient d'un auteur à l'autre; nous en resterons donc sur cet inventaire descriptif succinct ; la seule manière de poursuivre serait de procéder à des études monographiques auteur par auteur.

c) Découpage de syntagmes à l'intérieur d'un vers.

Ces contraintes s'ajoutent aux précédentes et sont de nature comparable : certains points intérieurs de chaque vers (hémistiche par exemple), importants du point de vue de la versification, doivent coïncider avec des charnières syntaxiques importantes. Par exemple, la charnière sujet -prédicat coïncidera souvent chez Racine avec l'hémistiche :

"Ma blessure trop vive / aussitôt a saigné" (Phèdre)
"Narcisse, plus hardi / s'empresse pour lui plaire" (Britannicus)
"Un faux Astyanax / fut offert au supplice" (Andromaque)
"Mon repos, mon bonheur / semblait être affermi" (Phèdre)

Parfois cette césure sépare le groupe verbal du complément d'objet dans la principale :

"Vous n'entreprenez point / une injuste carrière" (Bajazet, 441)

Charnière d'une coordination à l'hémistiche :

"Ma gloire me rappelle / et m'entraîne à l'autel" (Mithridate)
"L'autre, trop redoutable / et trop digne d'envie" (Bajazet,113)

d) D'autres contraintes.

Il n'est pas dans notre propos de donner une énumération complète des possibilités de contraintes mixtes faisant intervenir à la fois la versification et la syntaxe. Le caractère contraignant de l'ensemble est certain. Il n'est par exemple pas possible de trouver chez Racine, au niveau de l'hémistiche, une articulation peu importante telle que celle d'un déterminatif à un substantif. (Ce serait possible peut-être chez Victor Hugo, dans un alexandrin 4, 4, 4, mais avec alors des interdits en d'autres points du vers) ; et chez Racine, encore plus impossible de trouver une articulation déterminatif - nom à la charnière d'un vers.

V,2) Evaluation du niveau de complexité introduit par ces contraintes.

En termes de langages formels, certains niveaux de complexité peuvent être évalués en termes de "langages de Kleene", "C langage" etc...

Si les contraintes de versifications portaient uniquement sur les longueurs des phrases complètes, en supposant des contraintes syntaxiques du niveau de complexité d'un C langage, la résultante serait encore du niveau de complexité des C langages (intersection d'un K-langage et d'un C-langage).

Mais ce n'est pas le cas : les contraintes portent sur les différents syntagmes au fur et à mesure de leur assemblage pour constituer des phrases. Il n'est donc pas possible d'affirmer à priori que l'on reste dans le domaine des langages context free.

V,3) Conduite du calcul.

La conduite du calcul met en jeu un semi-anneau sur magma E avec produit externe par des pseudo scalaires αi élément d'un autre semi-anneau $\mathcal{E}$.

Soient $\{\gamma i\}$ les scalaires qui permettraient d'exprimer les contraintes syntaxiques considérées seules, et $\{\varphi i\}$ les scalaires qui permettraient d'exprimer les contraintes de versification. Les scalaires αi s'écriront comme des produits cartésiens :

$$\alpha i = \gamma i \times \varphi i$$

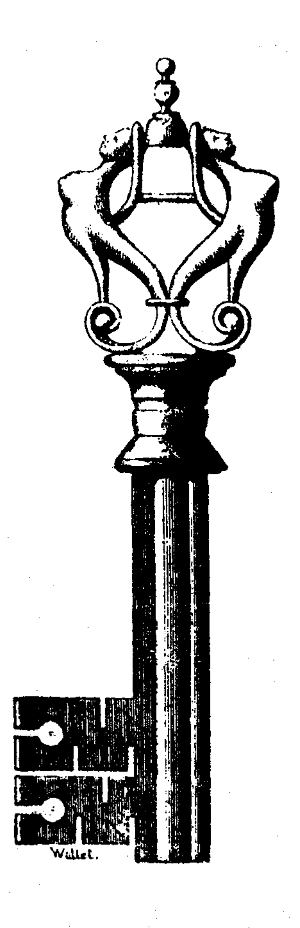
$$\alpha i \otimes \alpha j = (\gamma i, \varphi i) \otimes (\gamma j, \varphi j) = (\gamma k, \varphi k)$$

Par exemple, nous pouvons utiliser des paramètres φi associés à chaque mot du texte, qui soient caractéristiques de sa place dans la versification. La règle selon laquelle on ne trouve pas le couple "Art N" à cheval sur un hémistiche se formule comme suit :

$$(\text{Art}, \ {}_6\varphi_6) \ (\text{N}, \ {}_7\varphi_{7} + p) = 0$$

où ${}_6\varphi_6$ signifie que le mot considéré commence et finit sur la sixième syllabe du vers, et où ${}_7\varphi_{7+p}$ signifie que le mot (N) considéré commence à la septième syllabe du vers et finit sur la syllabe 7 + p.

La fonction de compilation est égale au produit contracté, comme dans tous les autres calculs analogues conduits plus haut.



AVERTISSEMENT

Le présent bulletin répond à une visée toute didactique : livrer sous forme accessible aux nouveaux venus dans les groupes de travail courants :

- de l'information technique et bibiographique en rapport avec leurs disciplines ;
- des programmes commentés de tous niveaux permettant un accès relativement rapide à des techniques de programmation appropriées, ainsi qu'une implémentation aisée.

On s'est efforcé, dans la mesure du possible, de ne pas établir de clivage trop net entre les disciplines artistiques et scientifiques concernées (musique, arts plastiques, poésie, architecture, logique, informatique),mais tout au contraire, ne serait-ce que par des techniques de programmation communes.

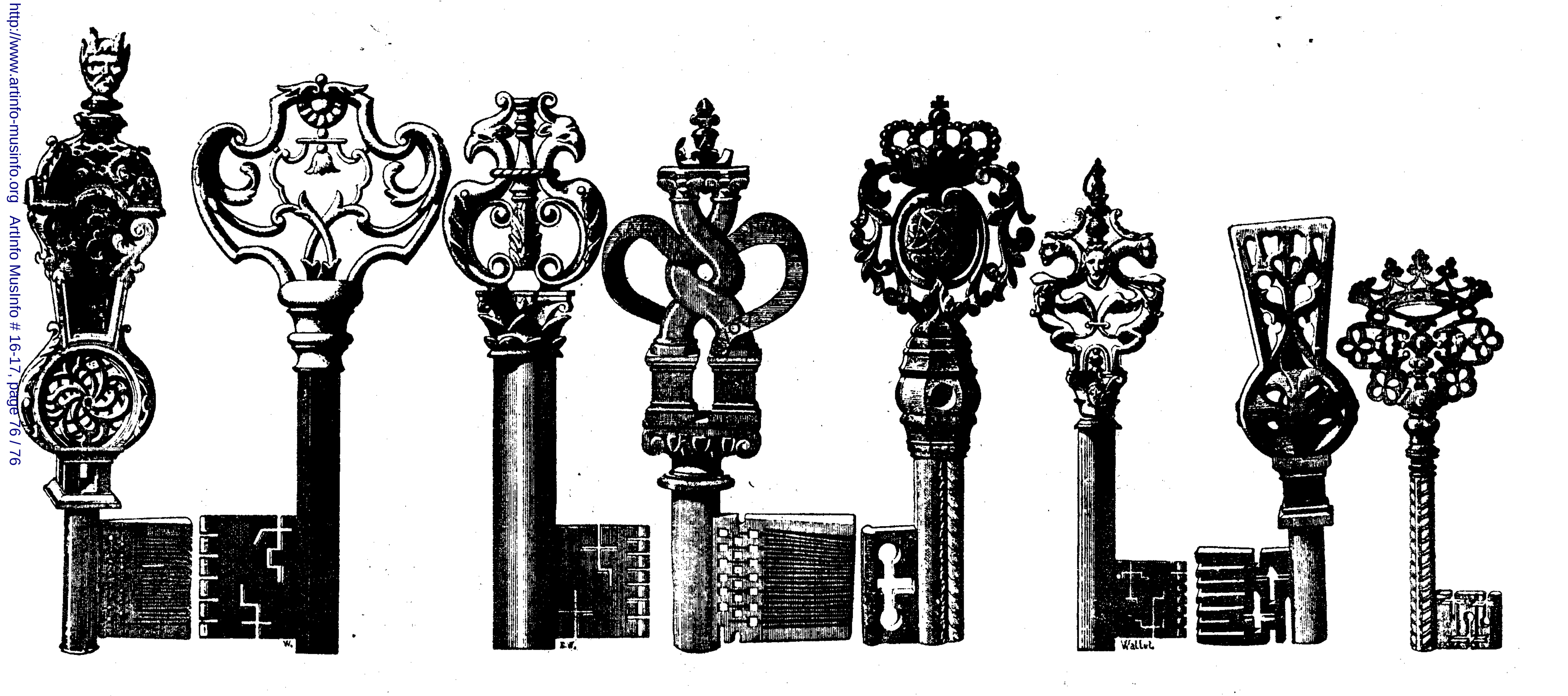
L'aspect pédagogique du présent bulletin reflète une préoccupation constante du groupe, à savoir ne pas se satisfaire en dernier ressort de méthodes de programmation trop élémentaires.

Le contenu des textes et des programmes n'engage que leurs auteurs.

Pour tous renseignements et composition des livraisons à venir,s'adresser à Jacques ARVEILLER, Département d'Informatique, Université Paris 8 Route de la Tourelle, Paris 12°.

Pour tout envoi, s'adresser à Patrick GREUSSAY, même adresse.

+=+=+=+=+=+=+=+=+



ARTINFO/MUSINFO

Université Paris 8

Institut d'Intelligence Artificielle

Groupe Art et Informatique

Bulletin Technique #16-17